STIC-Biotech/ChemLib

From: Seharaseyon, Jegatheesan
Sent: Tuesday, July 03, 2001 1:15 PM
To: STIC-Biotech/ChemLib
Subject: Re:09/284320

Please search SEQ ID No: 6 of 09/284,320. Thanks.

J.Seharaseyon
Art Unit 1647
CM1 10D16
10CO1 MB
(703)-305-1112

Access DB# _____

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # : _____ Date: _____
Art Unit: _____ Phone Number 30_____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

****

STAFF USE ONLY	Type of Search	Vendors and cost where applicable
Searcher: B Verner Jaly	NA Sequence (#) _____	STN _____
Searcher Phone #: 308-4501	AA Sequence (#) 1	Dialog _____
Searcher Location: Biotech Lib	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: 7/3/01	Bibliographic _____	Dr.Link _____
Date Completed: 7/3/01	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems: ABSSO2
Clerical Prep Time: 3 min	Patent Family _____	WWW/Internet _____
Online Time: 2 min	Other _____	Other (specify) _____

PTO-1590 (1-2000)

GenCore version 4.5

ON protein - protein search, using sw model

Run on: July 3, 2001, 16:15:33 ; Search time 22.94 Seconds
(without alignments)
924.952 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVFVVLLALVAGVLGNEFS.....MDPGVDSIIYRMTNQKIRMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1763	100.0	350	19	AAW64539	Human epidermoid c
2	1763	100.0	350	22	AAB90553	Human secreted pro
3	1757	99.7	350	22	AAB88347	Human membrane or
4	1751	99.3	350	22	AAB90589	Human secreted pro
5	1486	84.3	335	20	AAY59720	Secreted protein 3
6	441	25.0	93	20	AAY13118	Human secreted pro
7	121.5	6.9	364	21	AAG11429	Arabidopsis thalia
8	121.5	6.9	392	21	AAG11428	Arabidopsis thalia
9	103.5	5.9	428	19	AAW88378	H. pylori GHPO 137
10	101.5	5.8	318	21	AAG11430	Arabidopsis thalia
11	101	5.7	412	19	AAY10994	H. pylori ORF 09ce
12	100.5	5.7	1103	14	AAR39631	Neurofibromatosis
13	100.5	5.7	2485	15	AAR59921	RAS associated GAP
14	100.5	5.7	2485	15	AAR59922	RAS associated GAP
15	100.5	5.7	2818	13	AAR22268	Nf1 gene product.
16	100.5	5.7	2818	18	AAW13280	Human neurofibromi
17	97.5	5.5	982	21	AAR10105	Feline foamy virus
18	97	5.5	428	21	AAY28433	Corn leuC subunit
19	96.5	5.5	1886	19	AAW54341	Ratlus norvegicus
20	96	5.4	853	21	AAG47981	Arabidopsis thalia
21	96	5.4	916	21	AAG47980	Arabidopsis thalia
22	96	5.4	922	21	AAG47979	Human telomerase R
23	93	5.3	2627	19	AAW61347	Arabidopsis thalia
24	92	5.2	451	21	AAG40227	Arabidopsis thalia
25	92	5.2	469	21	AAG40226	Arabidopsis thalia
26	92	5.2	517	21	AAG40225	H. pylori GHPO 125
27	92	5.2	595	19	AAW98283	Chimeric Tsq DNA p
28	92	5.2	908	20	AAY33451	Arabidopsis thalia
29	91.5	5.2	454	21	AAG53794	Arabidopsis thalia
30	91.5	5.2	511	21	AAG53793	Arabidopsis thalia
31	91.5	5.2	545	21	AAG53792	Bacillus thuringie
32	91.5	5.2	789	19	AAW46868	Bacillus thuringie
33	91.5	5.2	789	20	AAY24971	A Streptococcus pn
34	91	5.2	448	21	AAY84831	A Streptococcus pn
35	91	5.2	457	21	AAY84830	Murine Ataxia-tela
36	91	5.2	3066	18	AAW36178	Aspergillus nidula
37	90.5	5.1	349	19	AAW44368	Arabidopsis thalia
38	90.5	5.1	438	21	AAG19277	Arabidopsis thalia
39	90.5	5.1	454	21	AAG19276	Arabidopsis thalia
40	90.5	5.1	454	21	AAG43053	Arabidopsis thalia
41	90.5	5.1	546	21	AAG43052	Membrane-bound pro
42	90.5	5.1	561	20	AAG43051	Human secreted pro
43	90.5	5.1	611	21	AAY66643	Amino acid sequenc
44	90.5	5.1	673	21	AAB38323	
45	90.5	5.1	673	21	AAB07428	

ALIGNMENTS

RESULT 1
AAW64539
ID AAW64539 standard; Protein; 350 AA.
XX
AC AAW64539;
XX
DT 21-OCT-1998 (first entry)
XX
DE Human epidermoid carcinoma cell line KB clone HP10034 protein.
XX
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
PN WO9821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX
PR 13-NOV-1996; 96JP-0301429.
XX
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
DR WPI; 1998-297932/26.
DR N-PSDB; AAV49560, AAV49561.
XX
PT Human protein having transmembrane domain - useful for, e.g.

PT research and nutrition
XX
PS Claim 1; Page 98-99; 205pp; English.
XX
CC AAW64534-W64558 represent human proteins containing a transmembrane
CC domain. These proteins can be used for, e.g. research and nutrition, and
CC may have cytokine and cell proliferation/differentiation, immune
CC stimulation/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
SQ Sequence 350 AA;

Query Match 100.0%; Score 1763; DB 19; Length 350;
Best Local Similarity 100.0%; Pred. No. 8.2e-165;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
Db 1 mavfvvllalvagvlgnefsilkspgsvvfrngnwpipgeripdvaalsmgfsvkedlsw 60

Qy 61 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEE 120
Db 61 pglavgnlfhrpratvmvmvkgvnklalppgsvisyplenavpfsldsvansihslfsee 120

Qy 121 TPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVD 180
Db 121 tpvvlqlapseervymvgkansvfedlsvtlrqlrnrlfqensvlsslplnslsrnnevd 180

Qy 181 LLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
Db 181 llflselqvlhdisslls rhkhlakdhspdlyslelaglde igkrygedseqfrdaskil 240

Qy 241 VDALQKFADDMYSLYGGNAVVELVTVKSPDTSLIRKTRTILEAKQAKNPASPYNLAYKYN 300
Db 241 vdalqkfaddmyslyggnavvelvtvksfdtslirktrtileakqaknpaspynlaykyn 300

Qy 301 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
Db 301 feysvvfnmvlwimialalavi itsyniwnmdpgydsiiyrmtnqkirmd 350

RESULT 2
AAB90553
ID AAB90553 standard; Protein; 350 AA.
XX
AC AAB90553;
XX
DT 01-JUN-2001 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 91.
XX
KW Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW nootropic; anticonvulsant; antialzheimers; antiparkinsonian;
KW protein coordinate data; infection.
XX
OS Homo sapiens.
XX
PN WO200121658-A1.
XX
PD 29-MAR-2001.
XX
PF 22-SEP-2000; 2000WO-US26013.
XX
PR 24-SEP-1999; 99US-0155709.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI Young PE, Wei P, Florence KA;
XX
DR WPI; 2001-235311/24.
DR N-PSDB; AAP97893.
XX
PT Nucleic acids encoding 32 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 11; Page 776-777; 890pp; English.
XX
CC The present sequence is one of 32 novel human secreted polypeptides. The
CC nucleic acid molecules and polypeptides may be used in the prevention,
CC diagnosis and treatment of diseases such as immune disorders (e.g.
CC multiple sclerosis, systemic lupus erythematosus and human
CC immuno-deficiency virus (HIV) infections), hyperproliferative disorders
CC (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC neovascularisation and diabetic retinopathy), neurological disorders
CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC infectious diseases and/or for promoting wound healing, regeneration
CC secreted polypeptides. They may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acid
CC sequences in samples. The polypeptides may be used as antigens in the
CC production of antibodies and in assays to identify modulators of
CC their expression and activity.
XX
SQ Sequence 350 AA;

Query Match 100.0%; Score 1763; DB 22; Length 350;
Best Local Similarity 100.0%; Pred. No. 8.2e-165;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
Db 1 mavfvvllalvagvlgnefsilkspgsvvfrngnwpipgeripdvaalsmgfsvkedlsw 60

Qy 61 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEE 120
Db 61 pglavgnlfhrpratvmvmvkgvnklalppgsvisyplenavpfsldsvansihslfsee 120

Qy 121 TPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVD 180
Db 121 tpvvlqlapseervymvgkansvfedlsvtlrqlrnrlfqensvlsslplnslsrnnevd 180

Qy 181 LLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
Db 181 llflselqvlhdisslls rhkhlakdhspdlyslelaglde igkrygedseqfrdaskil 240

Qy 241 VDALQKFADDMYSLYGGNAVVELVTVKSPDTSLIRKTRTILEAKQAKNPASPYNLAYKYN 300
Db 241 vdalqkfaddmyslyggnavvelvtvksfdtslirktrtileakqaknpaspynlaykyn 300

Qy 301 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
Db 301 feysvvfnmvlwimialalavi itsyniwnmdpgydsiiyrmtnqkirmd 350

RESULT 3
AAB88347
ID AAB88347 standard; Protein; 350 AA.
XX
AC AAB88347;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human membrane or secretory protein clone PSEC0072.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;

```
KW    rheumatoid arthritis; diabetes.
XX
OS    Homo sapiens.
XX
PN    EP1067182-A2.
XX
PD    10-JAN-2001.
XX
PF    07-JUL-2000; 2000EP-0114090.
XX
PR    08-JUL-1999;  99JP-0194179.
PR    11-JAN-2000;  2000JP-0018775.
PR    02-MAY-2000;  2000JP-0183766.
XX
PA    (HELI-) HELIX RES INST.
XX
PI    Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR    WPI; 2001-093989/11.
DR    N-PSDB; AAP93774.
XX
PT    Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT    gene therapy or as candidate target molecules in drug development -
PT
XX
PS    Claim 1; SEQ ID 62; 609pp + CD ROM; English.
XX
CC    This invention relates to nucleic acid sequences AAP93744 - AAP93916
CC    which encode human secretory or membrane proteins represented by
CC    AAB88317 - AAB88419. Included in the invention are primers
CC    AAP93917 - AAP94295 and AAP62232 - AAP62235 which are used to isolate the
CC    cDNA sequences of the invention. The invention also includes methods for
CC    the production of antibodies directed against the polynucleotide sequences
CC    sequences, which can be used in gene therapy. The polynucleotide sequences
CC    can be used in the prevention, treatment and
CC    proteins they encode may be used in appropriate secretory
CC    diagnosis of diseases associated with inappropriate acids and complementary
CC    protein/membrane protein expression. The nucleic acids and
CC    sequences may also be used as DNA probes in diagnostic assays
CC    presence of similar nucleic acid sequences in samples. They may also be
CC    (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC    used to study the expression and function of secretory proteins/membrane
CC    polypeptides and their role in metabolism. The polypeptides may be used
CC    as antigens in the production of antibodies against them and in assays to
CC    identify modulators (agonists and antagonists) and activity. The antibodies may also
CC    activity. The polypeptides may also be used as therapeutic
CC    agents to down regulate expression and function of
CC    be used as diagnostic agents for detecting the presence of
CC    polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC    (ELISA). Examples of diseases which may be treated include rheumatoid
CC    arthritis and diabetes.
XX
SQ    Sequence   350 AA;

  Query Match          99.7%;  Score 1757;  DB 22;  Length 350;
  Best Local Similarity  99.7%;  Pred. No. 3.2e-164;
  Matches  349;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAVFVWLLALVAGVLGNESILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 mavfvvllalvagvlgnefsilkspgsvvfrngnwpipgeripdvaalsmgfsvkedlsw 60

Qy         61 PGLAVGNLPHRPRATVHVMVKGVNKLALPPGSVISYPLENAVPPSLDSVANSIHSLFSEE 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 pglavgnlphrpratvmvmvkgvnklalppgsvisyplenavpfsldsvansihslfsee 120

Qy        121 TPVVLQLAPSEERVTWWGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVD 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 tpvvlqlapseervtwwgkansvfedlsvtlrqlrnrlfqensvlsslplnslsrnnevd 180

Qy        181 LLFLSEQVLHDISSLLSRHKHLAKDHSPDLYSLEAGLDEIGKRYGEDSEQFRDASKIL 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 llflseqvlhdissllsrhkhlakdhspdlyslealgldeigkrygedseqfrdaskil 240

Qy        241 VDALQKFADDMYSLVGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYN 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 vdalqkfaddmyslvggnavvelvtvksfdtslirktrtileakqaknpaspynlaykyn 300

Qy        301 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 feysvvfnmvlwimialalaviitsyniwnmdpgydsiiyrmtnqkirmd 350


RESULT   4
AAB90589  standard; protein; 350 AA.
ID
XX
AC    AAB90589;
XX
DT    01-JUN-2001  (first entry)
XX
DE    Human secreted protein, SEQ ID NO: 127.
XX
KW    Human; secreted protein; immunomodulatory; antiscleratic;
KW    dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW    vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW    nootropic; anticonvulsant; antialzheimers; antiparkinsonian;
KW    antimicrobial; vulnerary; vaccine; gene therapy; cancer;
KW    protein coordinate data; infection.
XX
OS    Homo sapiens.
XX
PN    WO200121658-A1.
XX
PD    29-MAR-2001.
XX
PF    22-SEP-2000; 2000WO-US26013.
XX
PR    24-SEP-1999;  99US-0155709.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI    Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI    Young PE, Wei P, Florence KA;
XX
DR    WPI; 2001-235311/24.
DR    N-PSDB; AAP97929.
XX
PT    Nucleic acids encoding 32 human secreted polypeptides, useful for
PT    preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT    disease and diabetic retinopathy -
PT
XX
PS    Claim 11; Page 809-810; 890pp; English.
XX
CC    The present sequence is one of 32 novel human secreted polypeptides. The
CC    nucleic acid molecules and polypeptides may be used in the prevention,
CC    diagnosis and treatment of diseases such as immune disorders (e.g.
CC    multiple sclerosis), systemic lupus erythematosus and human
CC    immuno-deficiency virus (HIV) infections), hyperproliferative disorders
CC    (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC    (e.g. Chagas's cardiomyopathy and coronary graft
CC    arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC    neovascularisation and diabetic retinopathy), neurological disorders
CC    (e.g. Huntington's chorea, Alsheimer's disease and Parkinson's disease),
CC    infectious diseases and/or for promoting wound healing, regeneration
CC    and/or chemotaxis. The nucleic acid molecules may be used to produce the
CC    secreted polypeptides. The nucleic acid molecules may be used as DNA probes in diagnostic
CC    assays to detect and quantitate the presence of similar nucleic acid
CC    sequences in samples. The polypeptides may be used as antigens in the
CC    production of antibodies and in assays to identify modulators of
CC    their expression and activity.
XX
SQ    Sequence   350 AA;
```

```
Query Match             99.3%;  Score 1751;  DB 22;  Length 350;
Best Local Similarity   99.4%;  Pred. No. 1.2e-163;
Matches  348;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 maxxvvllalvagvlgnefsilkspgsvvfrngnwpipgeripdvaalsmgfsvkedlsw 60

Qy         61 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 pglavgnlfhrpratvmvmvkgvnklalppgsvisyplenavpfsldsvansihslfsee 120

Qy        121 TPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 tpvvlqlapseervymvgkansvfedlsvtlrqlrnrlfqensvlsslplnslsrnnevd 180

Qy        181 LLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 llflselqvlhdissllsrhkhlakdhspdlyslelagldeigkrygedseqfrdaskil 240

Qy        241 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 vdalqkfaddmyslyggnavvelvtvksfdtslirktrtileakqaknpaspynlaykyn 300

Qy        301 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 feysvvfnmvlwimialalaviitsyniwnmdpgydsiiyrmtnqkirmd 350
```

RESULT 5
AAY59720
ID   AAY59720 standard; Protein; 335 AA.
XX
AC   AAY59720;
XX
DT   18-JAN-2000 (first entry)
XX
DE   Secreted protein 33-77-4-E2-FL1.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST) GENSET.
XX
PI   Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   N-PSDB; AAZ40848.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 10; Page 241; 244pp; English.
XX
CC   This sequence represents a human secreted protein of the invention.
CC   The extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence   335 AA;

```
Query Match             84.3%;  Score 1486;  DB 20;  Length 335;
Best Local Similarity   89.6%;  Pred. No. 1.3e-137;
Matches  302;  Conservative    7;  Mismatches   18;  Indels   10;  Gaps    3;

Qy          1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 mavfvvllalvagvlgnefsilkspgsvvfrngnwpipgeripdvaalsmgfsvkedlsw 60

Qy         61 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEE 120
              |||||||||||||:|||||||||    |||| ||||||||||||||||||||||||||||
Db         61 pglavgnlfhrprasvmvmvkgvnnxplppgcvisyplenavpfsldsvansihslfsee 120

Qy        121 TPVVLQLAPSEERVYMVGKANSV---FEDL--SVTLRQLRNRLFQENSVLSSLPLNSLSR 175
              |||||||||||||||||||||||   |:|  :  :  :  ||  :|      |  |||
Db        121 tpvvlqlapseervymvgkansvwktfqslapapxixcfkktlfsvhs-----pxixlsr 175

Qy        176 NNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRD 235
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 nnevdxlflselqvlhdissllsrhkhlakdhspdlyslelagldeigkrygedseqfrd 235

Qy        236 ASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNL 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 askilvdalqkfaddmyslyggnavvelvtvksfdtslirktrtileakqaknpaspynl 295

Qy        296 AYKYNFEYSVVFNMVLWIMIALALAVIITSYNIWNMD 332
              |||||||||||||||||||||||||||||||||||:
Db        296 aykynfeysvvfnmvlwimialalaviitsyniwnme 332
```

RESULT 6
AAY13118
ID   AAY13118 standard; Protein; 93 AA.
XX
AC   AAY13118;
XX
DT   22-JUN-1999 (first entry)
XX
DE   Human secreted protein encoded by 5' EST SEQ ID NO: 132.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS   Homo sapiens.
XX
PN   WO9906552-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01236.

```
XX
PR   01-AUG-1997;    97US-0905223.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153782/13.
DR   N-PSDB; AAX51918.
XX
PT   New isolated brain-derived nucleic acids - used to develop products
PT   which may have cytokine, immune, regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
XX
PS   Claim 34; Page 516; 577pp; English.
XX
CC   AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12987 to
CC   AAY13219, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence   93 AA;

  Query Match           25.0%;  Score 441;  DB 20;  Length 93;
  Best Local Similarity  95.6%;  Pred. No. 8.2e-36;
  Matches   86;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mavfvvllalvagvlgnefsilkspgsvvfrngnwpipgeripdvaalsmgfsvkedlsw 60

Qy      61 PGLAVGNLFHRPRATVMVMVKGVNKLALPP 90
           ||||||||||||||:|||||||||   |||
Db      61 pglavgnlfhrprasvmvmvkgvnxplpp 90


RESULT   7
AAG11429
ID   AAG11429 standard; Protein; 364 AA.
XX
AC   AAG11429;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 10142.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
```

```
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             6.9%;  Score 121.5;  DB 21;  Length 364;
  Best Local Similarity   22.0%;  Pred. No. 0.0017;
  Matches   86;  Conservative   72;  Mismatches  154;  Indels   79;  Gaps   18;

Qy       3 VFVVLLALVAGVLGNEFSILKSPGSVVF----------RNGNWPIPGERIPDVAALS-- 49
           |||| | |     |      :: ||| |           |: :  :|   :   ||:|
Db       6 vfvvallltlnyrg------easgsvffidgsnnqylrprssseal pmspveisaavsal 59

Qy      50 MGFSVKEDLSWPG------LAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVP 103
           :||:    |:  |      :   | | ||||  :: : | : : |      |: | ||:
Db      60 lgfapsatltadgssklnkilkpnpferpraafvleiagaddmlletspshsf-lgnair 118

Qy     104 FSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRL----- 158
            |: |        :  :|    :| :|   |  | : :|   |  :      : |
Db     119 ssiks------dsykadt----elpdnevvvvsvnepssdvtdkdi--ndfaswlggsyv 166

Qy     159 --FQENSVLSSLPLNSLSRNNEVDL-------LFLSELQVLHDISSLLSRHKHLAK--DH 207
              : :| | |:||     | | :|        |: | :  :|   :| :  |:    |
Db     167 agaepssgllsiplag-ganvefnlekeaerkfalnllglyqnirqavsvyddlshgidr 225

Qy     208 SPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVK 267
           : :|      |:| : : ||:     :    :|:  | |  : : : : |    ::| |
Db     226 taeltvgrfggidalaqeygqg--makqgmdvllstlsklfnlletshkg---qivgvi 279

Qy     268 SFDTSLIRKTRT1LEAKQAKNPA----------SPYNLAYKYNFEYSVVFNMVLWIM-IA 316
             |  : :::  :|    ::: |          |   :|     |  :|   :|:   |
Db     280 vldervnqesenllnfgssrssarsmvevegipsaaiia-----evilvrltlawltgii 334

Qy     317 LALAVIITSYNIWNMDPGYDSI1YRMTNQKI 347
           | :| |:  | : ||    |:::|  :| |:
Db     335 lliatilgvyflmnmpltkdtlly--snvkl 363


RESULT   8
AAG11428
ID   AAG11428 standard; Protein; 392 AA.
XX
AC   AAG11428;
XX
```

DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10141.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

```
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             6.9%;  Score 121.5;  DB 21;  Length 392;
  Best Local Similarity   22.0%;  Pred. No. 0.0019;
  Matches   86;  Conservative   72;  Mismatches  154;  Indels   79;  Gaps   18;

Qy      3 VFVVLLALVAGVLGNEFSILKSPGSVVF-----------RNGNWPIPGERIPDVAALS-- 49
            |||| | |      |        :: ||| |           |: :  :|   :   ||:|
Db     34 vfvvallltlnyrg------easgsvffidgsnnqylrprssseal pmspveisaavsal 87

Qy     50 MGFSVKEDLSWPG------LAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVP 103
           :||:     |:  |        :    | | |||| :: : | : : |      |: | ||:
Db     88 lgfapsatltadgssklnkilkpnpferpraafvleiagaddmlletspshsf-lgnair 146

Qy    104 FSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRL----- 158
            |: |          :  :|    :| :|  |  | : :|  |  :        : |
Db    147 ssiks------dsykadt----elpdnevvvvsvnepssdvtdkdi--ndfaswlggsyv 194

Qy    159 --FQENSVLSSLPLNSLSRNNEVDL-------LFLSELQVLHDISSLLSRHKHLAK--DH 207
              : :| | |:||   | | :|          |: | :  :|   :| :  |:  |
Db    195 agaepssgllsiplag-ganvefnlekeaerkfalnllglyqnirqavsvyddlshgidr 253

Qy    208 SPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVK 267
           : :|      |:| : : ||:     :    :|:  | |  : : : : |    ::| |
Db    254 taeltvgrfggidalaqeygqg--makqgmdvllstlsklfnlletshkg----qivgvi 307

Qy    268 SFDTSLIRKTRTILEAKQAKNPA----------SPYNLAYKYNFEYSVVFNMVLWIM-IA 316
               |  : :::  :|    ::: |               |    :|      |  :|    : |:  |
Db    308 vldervnqesenllnfgssrssarsmvevegipsaaiia-----evilvrltlawltgii 362

Qy    317 LALAVIITSYNIWNMDPGYDSIIYRMTNQKI 347
            | :| |:  | :  ||     |:::|  :| |:
Db    363 lliatilgvyflmnmpltkdtlly--snvkl 391


RESULT    9
AAW98378
ID   AAW98378 standard; Protein; 428 AA.
XX
AC   AAW98378;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 1375 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; AAX14097.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 8; Page 581-583; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   428 AA;


  Query Match             5.9%;  Score 103.5;  DB 19;  Length 428;
  Best Local Similarity   22.9%;  Pred. No. 0.13;
  Matches   72;  Conservative   35;  Mismatches  115;  Indels   93;  Gaps   15;

Qy     93 VISYPLENA------VPFSLDSVANSIHSLFSEETPVVLQLAPSEERV--------YMVG 138
            ||:  | ||         |    ||    | : | ||:        :|       :  :         |:
Db     21 viakellnaanrgvkvrilldd--ngldsdfsd-----imllnfhknievkifnpyyirn 73

Qy    139 KANSVFE---DLSVTLRQLRNRLFQENSVLSSLPLNSLSRN---NEVDLLFLSELQVL-- 190
            |     ||   |        ::: |:||  ::     :    ::   |    |::| ||  :|  |
Db     74 kglryfemladyerikkrmhnklfivdnfaviiggrnigdnyfdndldtnfl-dldalff 132
```

```
Qy    191 -----------------HDIS-SLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSE 231
                            | |||  || |  :          : |:| | |       ||
Db    133 ggvaskakesferywrfhrsipvsllrthkrlknn------akeiaklhekipisaedkn 186

Qy    232 QFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPAS 291
          ||        :|  ||:     | :| |||:    : |   || |    :    |  |  ||
Db    187 qfekkvndfidrfqky---qypiyygnaifladspkkidtplyspikiafe-kalknakd 242

Qy    292 PYNLAYKY----------------NFEYSVVFNMVLWIMIALALAVIITSYNIWNMDPG 334
             :|  |                  | ::: |        :|:     |   |   |
Db    243 svfiassyfipgkkmmkifknqiskgielniltn-------slsstdaivvygawe---- 291

Qy    335 YDSIIYRMTNQKIRM 349
                |  || :||
Db    292 ------ryrnqlvrm 300
```

RESULT 10
AAG11430
ID   AAG11430 standard; Protein; 318 AA.
XX
AC   AAG11430;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 10143.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144884.
PR   21-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145088.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145087.
PR   22-JUL-1999; 99US-0145089.
PR   22-JUL-1999; 99US-0145192.
PR   23-JUL-1999; 99US-0145145.
PR   23-JUL-1999; 99US-0145218.
PR   23-JUL-1999; 99US-0145224.
PR   26-JUL-1999; 99US-0145276.
PR   27-JUL-1999; 99US-0145913.
PR   27-JUL-1999; 99US-0145918.
PR   27-JUL-1999; 99US-0145919.
PR   28-JUL-1999; 99US-0145951.
PR   02-AUG-1999; 99US-0146386.
PR   02-AUG-1999; 99US-0146388.
PR   02-AUG-1999; 99US-0146389.
PR   03-AUG-1999; 99US-0147038.
PR   04-AUG-1999; 99US-0147204.

```
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
```

```
  Query Match             5.8%;  Score 101.5;  DB 21;  Length 318;
  Best Local Similarity   21.9%;  Pred. No. 0.13;
  Matches   75;  Conservative   66;  Mismatches  141;  Indels   61;  Gaps   16;

Qy      38 PGERIPDVAALSMGFSVKEDLSWPG------LAVGNLFHRPRATVMVMVKGVNKLALPPG 91
           | |    |:|| :||:     |: |         :    | | |||| :: : | : : |
Db       3 pveisaavsal-lgfapsatltadgssklnkilkpnpferpraafvleiagaddmllets 61

Qy      92 SVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTL 151
              |: | ||: |: |              :  :|     :|   :|   |  | : :|   | :
Db      62 pshsf-lgnairssiks------dsykadt----elpdnevvvvsvnepssdvtdkdi-- 108

Qy     152 RQLRNRL-------FQENSVLSSLPLNSLSRNNEVDL-------LFLSELQVLHDISSLL 197
               : |           : :| | |:||      | | :|              |: | :  :|   :
Db     109 ndfaswlggsyvagaepssgllsiplag-ganvefnlekeaerkfalnllglyqnirqav 167

Qy     198 SRHKHLAK--DHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLY 255
           | :  |:    | : :|        |:| : : ||:       :      :|:  | |  : : : :
Db     168 svyddlshgidrtaeltvgrfggidalaqeygqg--makqgmdvllstlsklfnlletsh 225

Qy     256 GGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPA----------SPYNLAYKYNFEYSV 305
               |     ::| |     |   : ::: :|      ::: |              |    :|        |  :
Db     226 kg----qivgvivldervnqesenllnfgssrssarsmvevegipsaaiia-----evil 276

Qy     306 VFNMVLWIM-IALALAVIITSYNIWNMDPGYDSIIYRMTNQKI 347
              |   : |:   | | :| |:   | : ||      |:::|   :| |:
Db     277 vrltlawltgiilliatilgvyflmnmpltkdtlly--snvkl 317
```

```
RESULT  11
AAY10994
ID   AAY10994 standard; Protein; 412 AA.
XX
AC   AAY10994;
XX
DT   08-JUN-1999  (first entry)
XX
DE   H. pylori ORF 09ce10413_35336707_f2_9 secreted protein.
XX
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
XX
OS   Helicobacter pylori.
XX
PN   WO9818323-A1.
XX
PD   07-MAY-1998.
XX
PF   28-OCT-1997;   97WO-US19575.
XX
PR   14-JUL-1997;   97US-0891928.
PR   28-OCT-1996;   96US-0739150.
PR   06-DEC-1996;   96US-0759739.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Alm RA,  Smith D;
XX
DR   WPI; 1998-271811/24.
DR   N-PSDB; AAX30461.
XX
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
XX
PS   Claims 27, 31; Page 201-202; 279pp; English.
XX
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
```

```
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
XX
SQ   Sequence   412 AA;

  Query Match             5.7%;  Score 101;  DB 19;  Length 412;
  Best Local Similarity   20.8%;  Pred. No. 0.21;
  Matches   69;  Conservative   50;  Mismatches   93;  Indels  120;  Gaps   17;

Qy      48 LSMGFSVKEDLSWPGLAVGNLFH-----------RPRATVMVMVKGVNKLALPPGSVI 94
             :|:| | | ||:    |     : :             : :  : :|: |:
Db      97 mslglsqkvdlngkkltqskminlekqkkilelkktkqqlvinlmingie--------- 146

Qy      95 SYPLENAVPF---SLDSVANSIHSLFSEETPVVLQLAPSE--ERVYMVGKANSVFEDLSV 149
             :|  :   :        :: :: |:::        :| :: :|  |  : :  : |       || |
Db     147 nyknqqeiellntaiknlentlyqanhssspdliaiakleilkslleiqk-----ndlev 201

Qy     150 TLR----QLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAK 205
             |        :        |:|| :|| | |     |||          | ||:||:
Db     202 alssshysmgeltfkeneilsiapkn-fefnne---------qelhnisa--------- 241

Qy     206 DHSPDLYSLELAGLDE---------IGKRYGEDSE-----QFRDASKILVDALQKFADDM 251
                 | : :| |||              | : ||          ||          : | :  ||
Db     242 ----tnydiaiarldeekaqkditlakksfledinvtgvyyfr--------skqyynydm 289

Qy     252 YS--------LYGGNAVVELVTVKSFDT------------------SLIRKTRTI---L 281
             :|           |||  |  :||  |  ::                  |::|  |:   |
Db     290 fsvalsiplplygkqa--klveqkkkeslafksevenaknktrhlalkllkkletlqknl 347

Qy     282 EA-----KQAKNPASPYNLAYKYNFEYSVVFN 308
             |:       || :  |  |  |  | | :|:  :|
Db     348 esinkiikqnekiaqiyaldlktngdynayyn 379

RESULT  12
AAR39631
ID   AAR39631 standard; Protein; 1103 AA.
XX
AC   AAR39631;
XX
DT   16-DEC-1993  (first entry)
XX
DE   Neurofibromatosis type 1 polypeptide.
XX
KW   Non-defective gene; NF-1; treatment; tumours; human; detection; ss.
XX
OS   Homo sapiens.
XX
PN   US5227292-A.
XX
PD   13-JUL-1993.
XX
PF   12-JUL-1990;   90US-0551531.
XX
PR   12-JUL-1990;   90US-0551531.
XX
PA   (UTAH ) UNIV UTAH.
XX
PI   Cawthon RM,  O'Connell P,  Viskochil DH,  White RL;
XX
DR   WPI; 1993-235118/29.
DR   N-PSDB; AAQ46263.
XX
PT   cDNA encoding neurofibromatosis type 1 gene - for detecting
PT   defective NF1 genes and tumours caused by such genes
XX
PS   Disclosure; Fig 7; 59pp; English.
XX
```

```
CC   The sequence is that of the neurofibromatosis type 1 (NF1)
CC   polypeptide which may be used therapeutically in the treatment of
CC   diseases associated with defective NF1 genes, e.g. tumours.
XX
SQ   Sequence   1103 AA;

  Query Match             5.7%;  Score 100.5;  DB 14;  Length 1103;
  Best Local Similarity   22.2%;  Pred. No. 1.1;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
             | ||:  : : :  :|   :||  :            :| |   ||    :::: | |  :|:
Db     416 gglgsikaevmadtavalasgnvklvsskvigrmckiidktclsptptleqhlmwddiai 475

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                      |  |:|:     | |  :              |::: |      | | ||  :     | | |||
Db     476 -------larymlmlsfnnsldvaahlpylfhvvtf-lvatgplslrasthglviniihs 527

Qy     116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
             |         |||||  ||:|: :|     : |::    : |         :    |    |:|  |     |
Db     528 lctcsqlhfseetkqvlrlsltefslpkfyllfgiskvksaaviafrssyrdrsfspgsy 587

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
              : | ||    |  || : |      : || :    | :     ||:       ::| |     |
Db     588 eretfaltsletvtea-lleime-acmrdiptckwldqwtelaqrfafqynpslqpralv 645

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
                 | ||  :|:  :  |  ||  |     |:      | |:
Db     646 vfgciskrvshgqikqiirilskalesclk--gpdtynsqvlieatvialtklqpllnkd 703

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                   :    ||:::|   | : |   |:|: :    |:: :   |  ||
Db     704 splhkalfwvavavlqldevnlysagtalleqnlhtldslrifndkspeevfmairnple 763

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                       | :   || :::| :::
Db     764 whckqmdhfvglnfnsnfnfalvghll 790

RESULT  13
AAR59921
ID   AAR59921 standard; protein; 2485 AA.
XX
AC   AAR59921;
XX
DT   22-FEB-1995  (first entry)
XX
DE   RAS associated GAP NF201.
XX
KW   Ras; GTPase activating protein; GAP; GAP related domain; GRD;
KW   pkT10; pkP11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;
KW   neurofibromatosis type 1; NF1.
XX
OS   Homo sapiens.
XX
PN   WO9416069-A.
XX
PD   21-JUL-1994.
XX
PF   12-JAN-1994;   94WO-US00198.
XX
PR   15-JAN-1993;   93US-0004824.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Kaziro Y,  Nakafuku M;
XX
DR   WPI; 1994-249216/30.
XX
PT   Blocking Ras-induced effects on a cell - by introducing a GTPase
PT   activating protein to the cell, used esp. in treatment of cancers
XX
```

```
PS   Disclosure; Page 36-44; 87pp; English.
XX
CC   Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
CC   cloned into the yeast expression vector pkT10 to obtain pkP11.  The
CC   pkP11 DNA was mutagenized by hydroxylamine in vitro and transformed
CC   into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
CC   RAS2Val19 mutation.  The heat shock sensitivity of the clones was
CC   checked.  Plasmid DNAs were recovered, re-transformed into TK161-
CC   R2V-D, and phenotypic reversion was examined.  2 Clones, NF201 and
CC   NF204 (given in AAR59922), which had strong suppression activity for
CC   RAS2Val19, were selected.  The mutant NF1-GRDs were also able to
CC   inhibit v-Ras-induced transformation in mammalian cells.
XX
SQ   Sequence   2485 AA;

  Query Match             5.7%;  Score 100.5;  DB 15;  Length 2485;
  Best Local Similarity   22.2%;  Pred. No. 3.9;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
           | ||:  : : :  :|   :||   :  :| |   ||   :::: | |  :|:
Db    1667 gglgsikaevmadtavalasgnvklvsskvigrmckiiidktclsptptleqhlmwddiai 1726

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
           |  |:|:  | |  :          |:::  |    | ||  :      | | |||
Db    1727 -------larymimlsfnnsldvaahlpylfhvvtf-lvatgplsirasthglvinilhs 1778

Qy     116 L-------PSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |       |||||  ||:|: :|    : |::   : |     :  |   |:| |    |
Db    1779 lctcsqlhfseetkqvlrlsltefslpkfyllfgiskvksaaviafrssyrdrsfspgsy 1838

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
            :  | ||     |  || : |   : || :   |  :   |:      ::| |    |
Db    1839 eretfaltsletvtea-lleime-acmrdiptckwldqwtelaqrfafqynpslqpralv 1896

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS---------------------- 253
           | ||  :|:  :  | ||  |    |:     | |:
Db    1897 vfgciskrvshgqikqiirilskalesclk--gpdtynsqvlieatvialtklqpllnkd 1954

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP------------ 292
            :       |||::|  | : |  |:|: :      |:: :  | ||
Db    1955 splhkalfwvavalqldevnlysagtalleqnlhtldslrifndkspeevfmairnple 2014

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                   | :  || :::| :::
Db    2015 whckqmdhfvglnfnsnfnfalvghll 2041


RESULT 14
AAR59922
ID   AAR59922 standard; protein; 2485 AA.
XX
AC   AAR59922;
XX
DT   22-FEB-1995  (first entry)
XX
DE   RAS associated GAP NF204.
XX
KW   Ras; GTPase activating protein; GAP; GAP related domain; GRD;
KW   pkP10; pkP11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;
KW   neurofibromatosis type 1; NF1.
XX
OS   Homo sapiens.
XX
PN   WO9416069-A.
XX
PD   21-JUL-1994.
XX
PF   12-JAN-1994;   94WO-US00198.
XX
PR   15-JAN-1993;   93US-0004824.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Kaziro Y,  Nakafuku M;
XX
DR   WPI; 1994-249216/30.
XX
PT   Blocking Ras-induced effects on a cell - by introducing a GTPase
PT   activating protein to the cell, used esp. in treatment of cancers
XX
PS   Disclosure; Page 44-52; 87pp; English.
XX
CC   Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
CC   cloned into the yeast expression vector pkT10 to obtain pkP11.  The
CC   pkP11 DNA was mutagenized by hydroxylamine in vitro and transformed
CC   into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
CC   RAS2Val19 mutation.  The heat shock sensitivity of the clones was
CC   checked.  Plasmid DNAs were recovered, re-transformed into TK161-
CC   R2V-D, and phenotypic reversion was examined.  2 Clones, NF201
CC   (given in AAR59921) and NF204, which had strong suppression activity
CC   for RAS2Val19, were selected.  The mutant NF1-GRDs were also able
CC   to inhibit v-Ras-induced transformation in mammalian cells.
XX
SQ   Sequence   2485 AA;

  Query Match             5.7%;  Score 100.5;  DB 15;  Length 2485;
  Best Local Similarity   22.2%;  Pred. No. 3.9;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
           | ||:  : : :  :|   :||   :  :| |   ||   :::: | |  :|:
Db    1667 gglgsikaevmadtavalasgnvklvsskvigrmckiiidktclsptptleqhlmwddiai 1726

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
           |  |:|:  | |  :          |:::  |    | ||  :      | | |||
Db    1727 -------larymimlsfnnsldvaahlpylfhvvtf-lvatgplsirasthglvinilhs 1778

Qy     116 L-------PSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |       |||||  ||:|: :|    : |::   : |     :  |   |:| |    |
Db    1779 lctcsqlhfseetkqvlrlsltefslpkfyllfgiskvksaaviafrssyrdrsfspgsy 1838

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
            :  | ||     |  || : |   : || :   |  :   |:      ::| |    |
Db    1839 eretfaltsletvtea-lleime-acmrdiptckwldqwtelaqrfafqynpslqpralv 1896

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS---------------------- 253
           | ||  :|:  :  | ||  |    |:     | |:
Db    1897 vfgciskrvshgqikqiirilskalesclk--gpdtynsqvlieatvialtklqpllnkd 1954

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP------------ 292
            :       |||::|  | : |  |:|: :      |:: :  | ||
Db    1955 splhkalfwvavalqldevnlysagtalleqnlhtldslrifndkspeevfmairnple 2014

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                   | :  || :::| :::
Db    2015 whckqmdhfvglnfnsnfnfalvghll 2041


RESULT 15
AAR22268
ID   AAR22268 standard; Protein; 2818 AA.
XX
AC   AAR22268;
XX
DT   06-MAY-1992  (first entry)
XX
DE   Nf1 gene product.
XX
KW   von Recklinghausen neurofibromatosis disease; autosomal dominant;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
```

PN WO9200387-A.
XX
PD 09-JAN-1992.
XX
PF 28-JUN-1991; 91WO-US04624.
XX
PR 29-JUN-1990; 90US-0547090.
XX
PA (UNMI) UNIV OF MICHIGAN.
XX
PI Collins FS, Wallace MR, Marchuk DA, Andersen LB, Gutmann DH;
XX
DR WPI; 1992-041568/05.
DR N-PSDB; AAQ20602.
XX
PT DNA sequences to von=Recklinghausen neurofibromatosis gene - and
PT derived amino acid sequences and probes for screening NF1 in early
PT stages of disease
XX
PS Claim 25; Page 67; 122pp; English.
XX
CC This is the amino acid sequence of the von Recklinghausen neuro-
CC fibromatosis (Nf1) gene product. It and antibodies raised to it
CC can be used in hybridisation and immunological assays to screen for
CC the presence of a normal or defective Nf1 gene product. Functional
CC assays to measure levels of gene function can also be used for
CC diagnosis or to monitor treatment. Patient therapy through
CC supplementation with the normal Nf1 product which can be
CC produced by recombinant techniques is also possible.
XX
SQ Sequence 2818 AA;

Query Match 5.7%; Score 100.5; DB 13; Length 2818;
Best Local Similarity 22.2%; Pred. No. 4.8;
Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

```
Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGPSVKEDLSWPGLAV 65
           | ||:  : : :  :|   :|| .     :| |    ||   :::: | |  :|:
Db    2000 qglgsikaevmadtavalasgnvklvsskvigrmcklidktclsptptleqhimwddiai 2059

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
             |  |:|:   | | :          |:::  |    | || :      | | |||
Db    2060 -------larymlmlsfnnsldvaahlpylfhvvtf-lvatgplslrasthglviniihs 2111

Qy     116 L-------FSEETPVVLQLAPSE---ERVTMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |       |||||  ||:|: :|    : |::   : |      :   |   |:| |    |
Db    2112 lctcsqlhfseetkqvlrlsltefslpkfyillfgiskvksaaviafrssyrdrsfspgsy 2171

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRWKHLAK----DHSPDLYSLELA 217
            :  | ||    |  || : |  : || :  | :   ||:    ::| |    |
Db    2172 eretfaltslsetvtea-lleime-acmrdiptckwldqwtelaqrfafqynpslqpralv 2229

Qy     218 GLDEIGKR--YGEDSEQPRDASKILVDALQKFADDMYS---------------------- 253
           ! ||  :|:  :  |  || |   |:    | |:
Db    2230 vfgciskrvshgqikqiirilskalesclk--gpdtynsqvlieatvialtklqpllnkd 2287

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP---------- 292
                :  ||::|  | : |  |:|: :    ::: :  |  ||
Db    2288 splhkalfwvavavlqldevnlysagtalleqnlhtldslrifndkspeevfmairnple 2347

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                   | :  || :::| :::
Db    2348 whckqmdhfvglnfnsnfnfalvghll 2374
```

Search completed: July 3, 2001, 16:18:21
Job time: 168 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:16:08 ; Search time 16.87 Seconds
(Without alignments)
417.938 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVFVVLLALVAGVLGNEPS.....MDPGYDSIIYRMTNQKIRMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	441	25.0	93	4	US-08-905-223-402	Sequence 402, App
2	100.5	5.7	2020	1	US-07-551-531-2	Sequence 2, Appli
3	100.5	5.7	2485	5	PCT-US94-00198-1	Sequence 1, Appli
4	100.5	5.7	2485	5	PCT-US94-00198-2	Sequence 2, Appli
5	100.5	5.7	2818	1	US-08-510-284-1	Sequence 1, Appli
6	100.5	5.7	2818	1	US-08-411-389-2	Sequence 2, Appli
7	100.5	5.7	2818	2	US-08-449-933-2	Sequence 2, Appli
8	100.5	5.7	2818	4	US-07-966-049A-2	Sequence 2, Appli
9	93	5.3	2627	2	US-08-751-189-3	Sequence 3, Appli
10	93	5.3	2627	2	US-09-060-836-3	Sequence 3, Appli
11	93	5.3	2627	4	US-09-184-445-3	Sequence 3, Appli
12	91	5.2	448	4	US-09-143-954-4	Sequence 4, Appli
13	91	5.2	457	4	US-09-143-954-2	Sequence 2, Appli
14	91	5.2	3066	4	US-08-952-127-12	Sequence 12, Appl
15	90.5	5.1	673	4	US-09-063-950-2	Sequence 2, Appli
16	90.5	5.1	2987	2	US-08-970-269A-29	Sequence 29, Appl
17	90.5	5.1	3959	2	US-08-970-269A-30	Sequence 30, Appl
18	89	5.0	903	2	US-08-209-521-24	Sequence 24, Appl
19	89	5.0	903	4	US-08-961-810-134	Sequence 134, App
20	89	5.0	903	4	US-08-352-902D-134	Sequence 134, App
21	88.5	5.0	790	4	US-08-960-780-4	Sequence 4, Appli
22	88.5	5.0	790	4	US-09-073-898-4	Sequence 4, Appli
23	88.5	5.0	1663	2	US-08-793-126-1	Sequence 1, Appli
24	88.5	5.0	1663	4	US-09-132-271-1	Sequence 1, Appli
25	88	5.0	473	5	PCT-US93-08386-9	Sequence 9, Appli
26	88	5.0	577	1	US-08-484-105-24	Sequence 24, Appl
27	88	5.0	577	1	US-08-484-106-24	Sequence 24, Appl
28	87.5	5.0	789	1	US-08-471-033-29	Sequence 29, Appl
29	87.5	5.0	789	2	US-08-471-044-29	Sequence 29, Appl
30	87.5	5.0	789	2	US-08-463-483A-29	Sequence 29, Appl
31	87.5	5.0	789	2	US-08-471-046A-29	Sequence 29, Appl
32	87.5	5.0	789	2	US-08-470-566B-29	Sequence 29, Appl
33	87.5	5.0	789	2	US-08-838-219B-2	Sequence 2, Appli
34	87.5	5.0	789	2	US-08-469-334-29	Sequence 29, Appl
35	87.5	5.0	789	3	US-09-300-529-29	Sequence 29, Appl
36	87.5	5.0	789	3	US-09-233-336A-2	Sequence 2, Appli
37	87.5	5.0	789	4	US-09-233-752A-2	Sequence 2, Appli
38	87.5	5.0	790	4	US-08-960-780-8	Sequence 8, Appli
39	87.5	5.0	790	4	US-09-073-898-8	Sequence 8, Appli
40	87.5	5.0	906	1	US-08-220-151-9	Sequence 9, Appli
41	87.5	5.0	906	1	US-08-413-118-9	Sequence 9, Appli
42	87.5	5.0	906	3	US-08-473-446-9	Sequence 9, Appli
43	87	4.9	336	3	US-09-095-163-2	Sequence 2, Appli
44	86.5	4.9	509	2	US-08-559-505-2	Sequence 2, Appli
45	86.5	4.9	509	2	US-08-890-980-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-905-223-402
; Sequence 402, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duelert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 402:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: -16..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 12.5
; OTHER INFORMATION: seq FVVLLALVAGVLG/NE
US-08-905-223-402

```
  Query Match            25.0%;  Score 441;  DB 4;  Length 93;
  Best Local Similarity   95.6%;  Pred. No. 3.9e-38;
  Matches   86;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60

Qy      61 PGLAVGNLFHRPRATVMVMVKGVNKLALPP 90
           ||||||||||||||:|||||||||   |||
Db      61 PGLAVGNLFHRPRASVMVMVKGVNNXPLPP 90


RESULT 2
US-07-551-531-2
; Sequence 2, Application US/07551531
; Patent No. 5227292
;  GENERAL INFORMATION:
;    APPLICANT:  WHITE, Raymond L.
;    APPLICANT:  O'CONNELL, Peter
;    APPLICANT:  VISKOCHIL, David H.
;    APPLICANT:  CAWTHON, Richard M.
;    TITLE OF INVENTION:  NEUROFIBROMATOSIS TYPE 1 GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/551,531
;      FILING DATE:  19900712
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  IHNEN, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  19780-94649
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2020 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-551-531-2

  Query Match            5.7%;  Score 100.5;  DB 1;  Length 2020;
  Best Local Similarity   22.2%;  Pred. No. 0.78;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
             | ||:  : : :  :|   :||  :            :| |   ||    :::: | |  :|:
Db    1202 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 1261

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                  |  |:|:   | |  :         |::: |    | ||  :        | | |||
Db    1262 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 1313

Qy     116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |          |||||  ||:|: :|     : |::    :        :    |:|  |     |
Db    1314 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 1373

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
                 :  | ||     | ||  :  |    : || :     |  :    ||:     ::|  |     |
Db    1374 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 1431

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
             | ||  :|:  :   |  ||  |   |:    | |:
Db    1432 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 1489

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                 :   ||::|  | :  |  |:|: :    |:: :   |   ||
Db    1490 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 1549

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                     | :   || :::| :::
Db    1550 WHCKQMDHFVGLNFNSNFNFALVGHLL 1576


RESULT 3
PCT-US94-00198-1
; Sequence 1, Application PC/TUS9400198
;  GENERAL INFORMATION:
;    APPLICANT:  Schering Corp.
;    TITLE OF INVENTION:  RAS Associated GAP Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  1 Girald Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  6.0.8
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00198
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/004,824
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Paul G.
;      REGISTRATION NUMBER:  32,743
;      REFERENCE/DOCKET NUMBER:  DX0352 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)822-7255
;      TELEFAX:  (201)822-7039
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2485 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  564..9380
PCT-US94-00198-1

  Query Match            5.7%;  Score 100.5;  DB 5;  Length 2485;
  Best Local Similarity   22.2%;  Pred. No. 1.1;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
             | ||:  : : :  :|   :||  :            :| |   ||    :::: | |  :|:
Db    1667 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 1726

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
```

```
              |  |:|:   | |  :         |::: |    | || :     | | |||
Db   1727 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 1778

Qy    116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
          |       |||||  ||:|: :|     : |::   : |      :  |    |:| |   |
Db   1779 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 1838

Qy    165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
            : | ||     |  || :  |  :  || :  |  :   ||:        ::| |   |
Db   1839 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 1896

Qy    218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
              | ||  :|:  :  |  || |   |:     | |:
Db   1897 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 1954

Qy    254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                 :   ||::|  | : |  |:|: :    |:: :  |  ||
Db   1955 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2014

Qy    293 ---------YNLAYKYNFEYSVVFNMV 310
                   | :  || :::| :::
Db   2015 WHCKQMDHFVGLNFNSNFNFALVGHLL 2041
```

RESULT 4
PCT-US94-00198-2
; Sequence 2, Application PC/TUS9400198
; GENERAL INFORMATION:
; APPLICANT: Schering Corp.
; TITLE OF INVENTION: RAS Associated GAP Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 1 Girald Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 6.0.8
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00198
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,824
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: DX0352 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)822-7255
; TELEFAX: (201)822-7039
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2485 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
PCT-US94-00198-2

Query Match 5.7%; Score 100.5; DB 5; Length 2485;
Best Local Similarity 22.2%; Pred. No. 1.1;
Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

```
Qy     13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
          | ||:  : : :  :|   :||  :           :| |    ||   :::: | |  :|:
Db   1667 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 1726

Qy     66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                 | |:|:   | | :           |::: |     | || :       | | |||
Db   1727 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 1778

Qy    116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
          |       |||||  ||:|: :|     : |::   : |      :  |    |:| |   |
Db   1779 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 1838

Qy    165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
            : | ||     |  || :  |  :  || :  |  :   ||:        ::| |   |
Db   1839 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 1896

Qy    218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
              | ||  :|:  :  |  || |   |:     | |:
Db   1897 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 1954

Qy    254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                 :   ||::|  | : |  |:|: :    |:: :  |  ||
Db   1955 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2014

Qy    293 ---------YNLAYKYNFEYSVVFNMV 310
                   | :  || :::| :::
Db   2015 WHCKQMDHFVGLNFNSNFNFALVGHLL 2041
```

RESULT 5
US-08-510-284-1
; Sequence 1, Application US/08510284
; Patent No. 5580955
; GENERAL INFORMATION:
; APPLICANT: Nur-E-Kamal, M. S. A.; Maruta, Hiroshi
; TITLE OF INVENTION: FRAGMENTS OF NEUROFIBROMIN (NF1) AND METHOD
; TITLE OF INVENTION: TO REVERSE ACTIVATED RAS INDUCED MALIGNANT TRANSFORMATION I
; TITLE OF INVENTION: MAMMALIAN CELLS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/510,284
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,575
; FILING DATE: 1-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5580955man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5319
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2818 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PUBLICATION INFORMATION:
; AUTHORS: Marchuk, Douglas A.; Saulino, Ann M.;

```
;      AUTHORS:  Tavakkol, Roxanne; Swaroop, Manju;
;      AUTHORS:  Wallace, Magaret R.; Andersen, Lone B.;
;      AUTHORS:  Mitchell, Anna L.; Gutmann, David H.;
;      AUTHORS:  Boguski, Mark; Collins, Francis S.
;      TITLE:  cDNA Cloning of the Type 1 Neurofibromatosis Gene:
;      TITLE:  Complete Sequence of the NF1 Gene Product
; Patent No. 5580955
;      JOURNAL:  Genomics
;      VOLUME:  11
;      PAGES:  931-940
;      DATE:  1991
US-08-510-284-1

  Query Match             5.7%;  Score 100.5;  DB 1;  Length 2818;
  Best Local Similarity   22.2%;  Pred. No. 1.3;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
             | ||:  : : :  :|   :||  :             :| |   ||   :::: | |  :|:
Db    2000 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 2059

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                     |  |:|:   | | :         |::: |    | || :     | | |||
Db    2060 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 2111

Qy     116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |        ||||| ||:|: :|     : |::   : |      :  |   |:| |   |
Db    2112 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 2171

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
              :  | ||     |  || : |    : || :    | :   ||:      ::| |    |
Db    2172 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 2229

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
                | ||   :|:   :   |  || |      |:      | |:
Db    2230 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 2287

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                   :    ||::|  | : |  |:|: :    |:: :   |  ||
Db    2288 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2347

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                    | :  || :::| :::
Db    2348 WHCKQMDHFVGLNFNSNFNFALVGHLL 2374

RESULT   6
US-08-411-389-2
; Sequence 2, Application US/08411389
; Patent No. 5605799
;  GENERAL INFORMATION:
;    APPLICANT:  White, Raymond L.
;    APPLICANT:  Cawthon, Richard M.
;    APPLICANT:  Li, Ying
;    TITLE OF INVENTION:  SOMATIC MUTATIONS IN THE
;    TITLE OF INVENTION:  NEUROFIBROMATOSIS TYPE 1 GENE IN HUMAN TUMORS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
;      STREET:  1201 New York Avenue NW, Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,389
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/047,088
;      FILING DATE:  16-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  19780-107116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2818 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-411-389-2

  Query Match             5.7%;  Score 100.5;  DB 1;  Length 2818;
  Best Local Similarity   22.2%;  Pred. No. 1.3;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
             | ||:  : : :  :|   :||  :             :| |   ||   :::: | |  :|:
Db    2000 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 2059

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                     |  |:|:   | | :         |::: |    | || :     | | |||
Db    2060 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 2111

Qy     116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |        ||||| ||:|: :|     : |::   : |      :  |   |:| |   |
Db    2112 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 2171

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
              :  | ||     |  || : |    : || :    | :   ||:      ::| |    |
Db    2172 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 2229

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
                | ||   :|:   :   |  || |      |:      | |:
Db    2230 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 2287

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                   :    ||::|  | : |  |:|: :    |:: :   |  ||
Db    2288 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2347

Qy     293 ---------YNLAYKYNFEYSVVFNMV 310
                    | :  || :::| :::
Db    2348 WHCKQMDHFVGLNFNSNFNFALVGHLL 2374

RESULT   7
US-08-449-933-2
; Sequence 2, Application US/08449933
; Patent No. 5859195
;  GENERAL INFORMATION:
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Wallace, Margaret R.
;    APPLICANT:  Marchuk, Douglas A.
;    APPLICANT:  Anderson, Lone B.
;    APPLICANT:  Guttman, David H.
;    TITLE OF INVENTION:  Neurofibromatosis Gene
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,933
;       FILING DATE:  25-MAY-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Konski, Antoinette F.
;       REGISTRATION NUMBER:  34,202
;       REFERENCE/DOCKET NUMBER:  20344-20553.10
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 813-5600
;       TELEFAX:  (415) 494-0792
;       TELEX:  706141 MRSNFOERS SFO
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2818 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:  17q11.2
;     FEATURE:
;       NAME/KEY:  Cleavage-site
;       LOCATION:  group(583..586, 815..818, 2573..2576, 2810..2813)
;       OTHER INFORMATION:  /note= "Potential cAMP-dependent
;       OTHER INFORMATION:  protein kinase recognition sites"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  2549..2556
;       OTHER INFORMATION:  /note= "Potential tyrosine
;       OTHER INFORMATION:  phosphorylation site"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  group(1264, 1276, 1358, 1377, 1389, 1390, 1391,
;       LOCATION:  1395, 1396, 1400, 1423, 1426, 1429, 1430)
;       OTHER INFORMATION:  /note= "Invariant residues within
;       OTHER INFORMATION:  most statistically significant regions of similarity among th
;       OTHER INFORMATION:  GAP family of proteins"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  group(1264..1290, 1345..1407, 1415..1430)
;       OTHER INFORMATION:  /note= "Most statistically
;       OTHER INFORMATION:  significant regions of similarity among the GAP family of
;       OTHER INFORMATION:  proteins"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  496
;       OTHER INFORMATION:  /note= "At variance with previously
;       OTHER INFORMATION:  published sequence which shows an ATG methionine codon rather
;       OTHER INFORMATION:  than an ATA isoleucine codon"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  1183
;       OTHER INFORMATION:  /note= "At variance with previously
;       OTHER INFORMATION:  published sequence. Shows an CTG leucine codon rather than th
;       OTHER INFORMATION:  previously published CTC"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  1555
;       OTHER INFORMATION:  /note= "At variance with previously
;       OTHER INFORMATION:  published sequence. Lacks an extra CAT histidine condon after
;       OTHER INFORMATION:  this residue"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  (2771^2772)
;       OTHER INFORMATION:  /note= "Position of an 18 amino
;       OTHER INFORMATION:  acid insertion(SEQ ID NO:10) representing an alternatively
;       OTHER INFORMATION:  spliced product"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  (1370^1371)
;       OTHER INFORMATION:  /note= "Position of a 21 amino acid
;       OTHER INFORMATION:  insertion representing an alternatively spliced product"
;     FEATURE:
;       NAME/KEY:  Domain
;       LOCATION:  1125..1537
;       OTHER INFORMATION:  /note= "NF1 catalytic domain"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  2746..2818
;       OTHER INFORMATION:  /note= "Corresponding amino acids
;       OTHER INFORMATION:  for the PstI-HindIII fragment designated pMAL.B3A"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  65..371
;       OTHER INFORMATION:  /note= "Corresponding amino acids
;       OTHER INFORMATION:  for the HpaI-PstI fragment designated pMAL.HF3A.P"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  65..1240
;       OTHER INFORMATION:  /note= "Corresponding amino acids
;       OTHER INFORMATION:  for the HpaI-XhoI fragment designated pMAL.HF3A.X"
;     PUBLICATION INFORMATION:
;       AUTHORS:  Wallace, M.R. et al.
;       TITLE:  Type 1 Neurofibromatosis Gene: Correction
;       JOURNAL:  Science
;       VOLUME:  250
;       ISSUE:  12/21/90
;       PAGES:  1749-
;       DATE:  12/21-1990
;       RELEVANT RESIDUES IN SEQ ID NO:  2: FROM 1 TO 2818
;     PUBLICATION INFORMATION:
;       AUTHORS:  Wallace, M.R. et al.
;       TITLE:  Type 1 Neurofibromatosis Gene: Identification
;       TITLE:  of a Large Transcript in Three NF1 Patients
;       JOURNAL:  Science
;       VOLUME:  249
;       ISSUE:  07/13/90
;       PAGES:  181-186
;       DATE:  07/13-1990
;       RELEVANT RESIDUES IN SEQ ID NO:  2: FROM 1 TO 2818
US-08-449-933-2

  Query Match            5.7%;  Score 100.5;  DB 2;  Length 2818;
  Best Local Similarity  22.2%;  Pred. No. 1.3;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
           | ||: : : :  :|   :||  :           :| |   ||   :::: | | :|:
Db    2000 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 2059

Qy      66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                   |  |:|:   | |  :         |::: |    | ||  :      | | |||
Db    2060 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 2111

Qy     116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
           |        |||||  ||:|: :|    | |::  : |    :  |   |:| |   |
Db    2112 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 2171

Qy     165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
              : | ||    | || : |   : || :   | :   ||:    ::| |    |
Db    2172 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 2229

Qy     218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
               | ||  :|:  :  |  || |   |:   | |:
Db    2230 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 2287

Qy     254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
```

```
                    :  ||::| | : |  |:|: :    |:: :  |  ||
Db    2288 SPLHKALFWVAVAVLQLDEVNLYSAGTALEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2347

Qy     293 ---------YNLAYKYNFETSVVFNMV 310
                    | :  || :::| :::
Db    2348 WHCKQMDHFVGLNFNSNFNFALVGLLL 2374


RESULT   8
US-07-966-049A-2
; Sequence 2, Application US/07966049A
; Patent No. 6238861
;  GENERAL INFORMATION:
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Wallace, Margaret R.
;    APPLICANT:  Marchuk, Douglas A.
;    APPLICANT:  Anderson, Lone B.
;    APPLICANT:  Guttman, David H.
;    TITLE OF INVENTION:  Neurofibromatosis Gene
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/966,049A
;      FILING DATE:  02-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Konski, Antoinette F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20344-20553.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141 MRSNFOERS SFO
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2818 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  17q11.2
;    FEATURE:
;      NAME/KEY:  Cleavage-site
;      LOCATION:  group(583..586, 815..818, 2573..2576, 2810..2813)
;      OTHER INFORMATION:  /note= "Potential cAMP-dependent
;      OTHER INFORMATION:  protein kinase recognition sites"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  2549..2556
;      OTHER INFORMATION:  /note= "Potential tyrosine
;      OTHER INFORMATION:  phosphorylation site"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  group(1264, 1276, 1358, 1377, 1389, 1390, 1391,
;      LOCATION:  1395, 1396, 1400, 1423, 1426, 1429, 1430)
;      OTHER INFORMATION:  /note= "Invariant residues within
;      OTHER INFORMATION:  most statistically significant regions of similarity among th
;      OTHER INFORMATION:  GAP family of proteins"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  group(1264..1290, 1345..1407, 1415..1430)
;      OTHER INFORMATION:  /note= "Most statistically
;      OTHER INFORMATION:  significant regions of similarity among the GAP family of
;      OTHER INFORMATION:  proteins"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  496
;      OTHER INFORMATION:  /note= "At variance with previously
;      OTHER INFORMATION:  published sequence which shows an ATG methionine codon rat
;      OTHER INFORMATION:  than an ATA isoleucine codon"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1183
;      OTHER INFORMATION:  /note= "At variance with previously
;      OTHER INFORMATION:  published sequence. Shows an CTG leucine codon rather than
;      OTHER INFORMATION:  previously published CTC"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1555
;      OTHER INFORMATION:  /note= "At variance with previously
;      OTHER INFORMATION:  published sequence. Lacks an extra CAT histidine condon af
;      OTHER INFORMATION:  this residue"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  (2771-2772)
;      OTHER INFORMATION:  /note= "Position of an 18 amino
;      OTHER INFORMATION:  acid insertion(SEQ ID NO:10) representing an alternatively
;      OTHER INFORMATION:  spliced product"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  (1370-1371)
;      OTHER INFORMATION:  /note= "Position of a 21 amino acid
;      OTHER INFORMATION:  insertion representing an alternatively spliced product"
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  1125..1537
;      OTHER INFORMATION:  /note= "NF1 catalytic domain"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  2746..2818
;      OTHER INFORMATION:  /note= "Corresponding amino acids
;      OTHER INFORMATION:  for the PstI-HindIII fragment designated pMAL.B3A"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  65..371
;      OTHER INFORMATION:  /note= "Corresponding amino acids
;      OTHER INFORMATION:  for the HpaI-PstI fragment designated pMAL.HF3A.P"
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  65..1240
;      OTHER INFORMATION:  /note= "Corresponding amino acids
;      OTHER INFORMATION:  for the HpaI-XhoI fragment designated pMAL.HF3A.X"
;    PUBLICATION INFORMATION:
;      AUTHORS:  Wallace, M.R. et al.
;      TITLE:  Type 1 Neurofibromatosis Gene: Correction
;      JOURNAL:  Science
;      VOLUME:  250
;      ISSUE:  12/21/90
;      PAGES:  1749-
;      DATE:  12/21-1990
;      RELEVANT RESIDUES IN SEQ ID NO:  2: FROM 1 TO 2818
;    PUBLICATION INFORMATION:
;      AUTHORS:  Wallace, M.R. et al.
;      TITLE:  Type 1 Neurofibromatosis Gene: Identification
;      TITLE:  of a Large Transcript in Three NF1 Patients
;      JOURNAL:  Science
;      VOLUME:  249
;      ISSUE:  07/13/90
;      PAGES:  181-186
;      DATE:  07/13-1990
```

```
;      RELEVANT RESIDUES IN SEQ ID NO:  2: FROM 1 TO 2818
US-07-966-049A-2

  Query Match             5.7%;  Score 100.5;  DB 4;  Length 2818;
  Best Local Similarity   22.2%;  Pred. No. 1.3;
  Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy       13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
               | ||:  : : :  :|   :||  :               :| |    ||    :::: | |   :|:
Db     2000 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 2059

Qy       66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                               |  |:|:    | |  :             |::: |     | ||  :           | |  |||
Db     2060 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 2111

Qy      116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
              |               |||||  ||:|: :|      : |::     :  |        :    |      |:| |     |
Db     2112 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 2171

Qy      165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
                    :  | ||      |   || : |    : || :      | :     ||:        ::| |      |
Db     2172 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 2229

Qy      218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS---------------------- 253
                   | ||   :|:   :    |  || |     |:       | |:
Db     2230 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 2287

Qy      254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP------------ 292
                                 :     ||::|   | : |   |:|: :      |:: :   |   ||
Db     2288 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2347

Qy      293 ---------YNLAYKYNFEYSVVFNMV 310
                                      |  :   || :::| :::
Db     2348 WHCKQMDHFVGLNFNSNFNFALVGHLL 2374
```

RESULT 9
US-08-751-189-3
; Sequence 3, Application US/08751189
; Patent No. 5919656
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,189
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2627 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-751-189-3

```
  Query Match             5.3%;  Score 93;  DB 2;  Length 2627;
  Best Local Similarity   22.5%;  Pred. No. 7.1;
  Matches   66;  Conservative   37;  Mismatches  112;  Indels   78;  Gaps   10;

Qy       22 LKSPGSV--VFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVM 79
               || ||::    :|:   | :      :|   |                | |                |  | :::
Db     1226 LKEPGALPSTYRSLVWELQQRLLPKSA---------ESL------------HPGQTQVLI 1264

Qy       80 VKGVNKLALPPGSVISYPLENAVP------FSLDSVANSIHSLFSEETPVVLQLAPSEER 133
               : |  ::|    | :||  :    :|        |: |  |     :|     :    || | | |
Db     1265 IDGADRLVDQNGQLISDWIPKKLPRCVHLVLSVSSDAGLGETLEQSQGAHVLALGPLEAS 1324

Qy      134 VYMVGKANSVFEDLSVTLRQLRNRLFQENSVL------SSLPLNSLSRNNEVDLLFLSE- 186
                    :|  | |:|::  ::|     |      |       | ||         : : |  | |
Db     1325 ----ARARLVREELALYGKRLEESPFNNQMRLLLVKRESGRPLYLRLVTDHLRLFTLYEQ 1380

Qy      187 ----LQVLHDISSLLSRH--KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
                        |: |      || :|     | |:| ||:      |   | |: :
Db     1381 VSERLRTLPATVPLLLQHILSTLEKEHGPDVLPQALTAL-EVTR--------------- 1423

Qy      241 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPY 293
                                         |  | :|  | |   :| :  |::  ||  |  |    ||
Db     1424 --------------SGLTVDQLHGVLSVWRTLPKGTKSWEEAVAAGNSGDPY 1461
```

RESULT 10
US-09-060-836-3
; Sequence 3, Application US/09060836
; Patent No. 5981707
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,836
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,189
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2627 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-060-836-3

Query Match 5.3%; Score 93; DB 2; Length 2627;

```
  Best Local Similarity   22.5%;  Pred. No. 7.1;
  Matches   66;  Conservative   37;  Mismatches  112;  Indels   78;  Gaps   10;

Qy      22 LKSPGSV--VFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVM 79
              || ||::    :|:  | :     :|  |           | |              |  | :::
Db    1226 LKEPGALPSTYRSLVWELQQRLLPKSA---------ESL-----------HPGQTQVLI 1264

Qy      80 VKGVNKLALPPGSVISYPLENAVP------FSLDSVANSIHSLFSEETPVVLQLAPSEER 133
              : | ::|    | :||  :    :|         |: | |     :|    :    || | | |
Db    1265 IDGADRLVDQNGQLISDWIPKKLPRCVHLVLSVSSDAGLGETLEQSQGAHVLALGPLEAS 1324

Qy     134 VYMVGKANSVFEDLSVTLRQLRNRLFQENSVL------SSLPLNSLSRNNEVDLLFLSE- 186
                :|  | |:|::  ::|      |          |        | ||       : :  |  ||
Db    1325 ----ARARLVREELALYGKRLEESPFNNQMRLLLVKRESGRPLYLRLVTDHLRLFTLYEQ 1380

Qy     187 ----LQVLHDISSLLSRH--KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
                |: |       || :|      | |:| ||:        |    | |: :
Db    1381 VSERLRTLPATVPLLLQHILSTLEKEHGPDVLPQALTAL-EVTR--------------- 1423

Qy     241 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPY 293
                                   |   | :|  | |     :|  :  |::   ||   |  |     ||
Db    1424 --------------SGLTVDQLHGVLSVWRTLPKGTKSWEEAVAAGNSGDPY 1461
```

RESULT 11
US-09-184-445-3
; Sequence 3, Application US/09184445
; Patent No. 6174703
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/184,445
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,189
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2627 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-184-445-3

```
  Query Match             5.3%;  Score 93;  DB 4;  Length 2627;
  Best Local Similarity   22.5%;  Pred. No. 7.1;
  Matches   66;  Conservative   37;  Mismatches  112;  Indels   78;  Gaps   10;

Qy      22 LKSPGSV--VFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVM 79
              || ||::    :|:  | :     :|  |           | |              |  | :::
Db    1226 LKEPGALPSTYRSLVWELQQRLLPKSA---------ESL-----------HPGQTQVLI 1264

Qy      80 VKGVNKLALPPGSVISYPLENAVP------FSLDSVANSIHSLFSEETPVVLQLAPSEER 133
              : | ::|    | :||  :    :|         |: | |     :|    :    || | | |
Db    1265 IDGADRLVDQNGQLISDWIPKKLPRCVHLVLSVSSDAGLGETLEQSQGAHVLALGPLEAS 1324

Qy     134 VYMVGKANSVFEDLSVTLRQLRNRLFQENSVL------SSLPLNSLSRNNEVDLLFLSE- 186
                :|  | |:|::  ::|      |          |        | ||       : :  |  ||
Db    1325 ----ARARLVREELALYGKRLEESPFNNQMRLLLVKRESGRPLYLRLVTDHLRLFTLYEQ 1380

Qy     187 ----LQVLHDISSLLSRH--KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
                |: |       || :|      | |:| ||:        |    | |: :
Db    1381 VSERLRTLPATVPLLLQHILSTLEKEHGPDVLPQALTAL-EVTR--------------- 1423

Qy     241 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPY 293
                                   |   | :|  | |     :|  :  |::   ||   |  |     ||
Db    1424 --------------SGLTVDQLHGVLSVWRTLPKGTKSWEEAVAAGNSGDPY 1461
```

RESULT 12
US-09-143-954-4
; Sequence 4, Application US/09143954A
; Patent No. 6194170
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: MurF
; FILE REFERENCE: GM10097
; CURRENT APPLICATION NUMBER: US/09/143,954A
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 60/060,011
; EARLIER FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-143-954-4

```
  Query Match             5.2%;  Score 91;  DB 4;  Length 448;
  Best Local Similarity   22.5%;  Pred. No. 0.61;
  Matches   48;  Conservative   39;  Mismatches   66;  Indels   60;  Gaps   11;

Qy      98 LENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNR 157
              || |:    :      |:  ::: :    :  | ||     |||:::            |: |  :|     |||
Db     252 LEQALDLPVTGKYNATNAMIA--SYVALQEGVSEEQIH------QAFQGLELT----RNR 299

Qy     158 --------------LFQEN--------SVLSSLPLNSLSRNNEVDLLFLSELQVLHDIS 194
                                 :: |              |::| |     :         :   |:::: | | |
Db     300 TEWKKAANGADILSDVYNANPTAMKLILETFSAIPANEGGKK----IAVLADMKELGDQS 355

Qy     195 SLLSRHKHLAKDHSPDL------YSLELAGLDE-------IGKRY----GEDSEQFRDAS 237
                 :   |    :        |||:           |   ::| | :            || |           || :| | |
Db     356 --IQLHNQMILSLSPDVLDTVIFYGEDIAELAQLASQMFPIGHVYYFKKTEDQDQFEDLV 413

Qy     238 KILVDALQKFADDMYSLYGGNAVVELVTVKSFD 270
              | : ::|     |:|     | |  |::     : |:| :
Db     414 KQVKESLS--ANDQILLKGSNSMNLAMLVESLE 444
```

RESULT 13
US-09-143-954-2
; Sequence 2, Application US/09143954A
; Patent No. 6194170
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: MurF
; FILE REFERENCE: GM10097
; CURRENT APPLICATION NUMBER: US/09/143,954A
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 60/060,011
; EARLIER FILING DATE: 1997-09-25

```
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 457
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-143-954-2

  Query Match             5.2%;  Score 91;  DB 4;  Length 457;
  Best Local Similarity   22.5%;  Pred. No. 0.63;
  Matches   48;  Conservative   39;  Mismatches   66;  Indels   60;  Gaps   11;

Qy      98 LENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNR 157
              || |:    :     |: ::: :   : | ||     |||:::          |:  |  :|      |||
Db     261 LEQALDLPVTGKYNATNAMIA--SYVALQEGVSEEQIH------QAFQGLELT----RNR 308

Qy     158 --------------LFQEN--------SVLSSLPLNSLSRNNEVDLLFLSELQVLHDIS 194
                             ::  |              |::|  |     :       :   |::::  | | |
Db     309 TEWKKAANGADILSDVYNANPTAMKLILETFSAIPANEGGKK----IAVLADMKELGDQS 364

Qy     195 SLLSRHKHLAKDHSPDL------YSLELAGLDE-------IGKRY----GEDSEQFRDAS 237
               :   |   :      |||:          |   ::|  |  :          ||   |         || :|| |
Db     365 --IQLHNQMILSLSPDVLDTVIFYGEDIAELAQLASQMFPIGHVYYFKKTEDQDQFEDLV 422

Qy     238 KILVDALQKFADDMYSLYGGNAVVELVTVKSFD 270
              |  :  ::|    |:|    |  |  |::    :  |:|  :
Db     423 KQVKESLS--ANDQILLKGSNSMNLAMLVESLE 453
```

RESULT 14
US-08-952-127-12
; Sequence 12, Application US/08952127
; Patent No. 6211336
; GENERAL INFORMATION:
;   APPLICANT: Shiloh, Yosef
;   APPLICANT: Tagle, Danilo A.
;   APPLICANT: Collins, Francis S.
;   TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Kohn & Associates
;   STREET: 30500 No. 6211336thwestern Hwy., Suite 410
;   CITY: Farmington Hills
;   STATE: Michigan
;   COUNTRY: U.S.
;   ZIP: 48334
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/952,127
;   FILING DATE:
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Kohn, Kenneth I.
;   REGISTRATION NUMBER: 30,995
;   REFERENCE/DOCKET NUMBER: 2290-00029
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 810-539-5050
;   TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3066 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;   ORGANISM: Mus musculus
US-08-952-127-12

```
  Query Match             5.2%;  Score 91;  DB 4;  Length 3066;
  Best Local Similarity   23.1%;  Pred. No. 15;
  Matches   50;  Conservative   41;  Mismatches   97;  Indels   28;  Gaps    7;

Qy     139 KANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLS 198
               :  |   |  |     |: ||||  |    ::      ||      || |   ||: |        |: :: :     ||
Db    2126 EGTSYHESLYNALQCLRNREF--STFYESLRYASLFRVKEVEELSKGSLESVYSLYPTLS 2183

Qy     199 RHKHLAK-DHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGG 257
              | : : : ::| :|:|  :             :   |       :::    |::|  |:     | : : :|
Db    2184 RLQAIGELENSGELFSRSVT-----DRERSEAYWKWQKHSQLLKDSDFSFQEPLMAL--R 2236

Qy     258 NAVVELVTVKSFDTSLIRKTRTILE---------AKQAKNPASPYNLAYK---YN----- 300
                ::| :   |   :  |       :: ||                |:   ||     |       :|     ||
Db    2237 TVILETLVQKEMERSQGACSKDILTKHLVEFSVLARTFKNTQLPERAIFKIKQYNSAICG 2296

Qy     301 -FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGY 335
              |: :        | |          :||: |        |   :|    :
Db    2297 ISEWHLEEAQVFWAKKEQSLALSILKQMIKKLDSSF 2332
```

RESULT 15
US-09-063-950-2
; Sequence 2, Application US/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: MEI-019
; CURRENT APPLICATION NUMBER: US/09/063,950C
; CURRENT FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 673
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-063-950-2

```
  Query Match             5.1%;  Score 90.5;  DB 4;  Length 673;
  Best Local Similarity   24.6%;  Pred. No. 1.4;
  Matches   75;  Conservative   39;  Mismatches  108;  Indels   83;  Gaps   16;

Qy       5 VVLLALVAGVLGNEFSILKSPGSVVF---RNGNWPIPGERIPDVAALSM---GFSVKEDL 58
             ::||||  || |             |        ||    | |    :| :   ||     | :     | :: :
Db      12 LLLLALGPGVQGCPSGCQCSQPQTVFCTARQGT-TVPRDVPPDTVGLYVFENGITMLDAG 70

Qy      59 SWPGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFS 118
             |: ||                 : ::       |::|   |    |   || |     :||  || :| : :
Db      71 SFAGL----------PGLQLLDLSQNQIASLPSGVFQ-PLANL--SNLDLTANRLHEITN 117

Qy     119 EETPVVLQLAPSEERVYMVGKAN------SVFEDLSVTLRQLRNRLFQENSVLSSLPLN- 171
             |       : :|      ||:|: ||         |:       || :|      |:| :  :   ||
Db     118 ETFRGLRRL----ERLYL-GKNRIRHIQPGAFD----TLDRLLELKLQDNELRALPPLRL 168

Qy     172 ------SLSRNN-----------EVDLLFLSELQVLHDISSLLSRHKHLAKDHSPD--- 210
                      || |:                  |: | |: | :        | || ::|      |    |
Db     169 PRLLLLDLSHNSLLALEPGILDTANVEALRLAGLGLQQLDEGLFSRLRNL---HDLDVSD 225

Qy     211 --------------LYSLELAGLDEIGKRYGED-----SEQFRDASKILVDALQKFADD 250
                             |  | |||   | :    ||         : |   | |  :      :||      |
Db     226 NQLERVPPVIRGLRGLTRLRLAGNTRIAQLRPEDLAGLAALQELDVSNL---SLQALPGD 282

Qy     251 MYSLY 255
               :  |:
Db     283 LSGLF 287
```

Search completed: July 3, 2001, 16:18:53
Job time: 165 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:16:43 ; Search time 16.08 Seconds
(without alignments)
1658.028 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVFVVLLALVAGVLGNEFS.....MDPGYDSIIYRMTNQKIRMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1667	94.6	330	2	T08667	hypothetical prote
2	192.5	10.9	324	2	T23876	hypothetical prote
3	121.5	6.9	364	2	T52136	probable type 1 me
4	114	6.5	663	2	T40039	hypothetical prote
5	109.5	6.2	408	2	T25243	hypothetical prote
6	105	6.0	450	2	S35732	dnaA protein - Spi
7	105	6.0	458	2	T19941	hypothetical prote
8	104	5.9	412	2	G64685	hypothetical prote
9	103.5	5.9	502	2	F64543	conserved hypothet
10	103.5	5.9	732	2	T43619	secreted protein k
11	103.5	5.9	732	2	S30060	protein kinase ypk
12	103.5	5.9	925	2	C84538	probable LRR recep
13	103	5.8	470	2	A71685	guanosine pentapho
14	103	5.8	979	2	JQ0894	P115 protein - Myc
15	102.5	5.8	1004	2	A39611	probable GTP-bindi
16	101	5.7	412	2	B71831	hypothetical prote
17	101	5.7	689	2	T40364	hypothetical prote
18	100.5	5.7	1358	2	A29360	SIR4 protein - yea
19	100.5	5.7	2818	2	B55282	neurofibromatosis-
20	100.5	5.7	2820	2	JC5196	neurofibromin I -
21	99	5.6	524	2	S77851	probable aspartate
22	98.5	5.6	377	2	S21302	succinate dehydrog
23	98.5	5.6	471	2	T48743	probable 26S ATP/u
24	98.5	5.6	1038	1	MWRBCB	myosin beta heavy
25	98.5	5.6	1935	1	A37102	myosin beta heavy
26	98	5.6	561	2	F75191	hypothetical prote
27	98	5.6	864	2	T08575	protein kinase hom
28	97.5	5.5	758	2	H75013	hypothetical prote
29	97.5	5.5	884	2	S73302	preprotein translo
30	97.5	5.5	2825	2	I54352	neurofibromin - mo
31	96.5	5.5	533	2	G72593	hypothetical prote
32	96.5	5.5	574	2	T51799	CLB1-like protein
33	96.5	5.5	1062	2	F83335	RND multidrug effl
34	96.5	5.5	1062	2	T30830	hypothetical prote
35	96.5	5.5	1934	2	T48153	myosin heavy chain
36	96.5	5.5	1938	1	S06005	myosin alpha heavy
37	96	5.4	390	2	E81122	bacteriophage DNA
38	95.5	5.4	1020	2	B82427	sensor protein Tor
39	95.5	5.4	1209	2	T21455	hypothetical prote
40	95.5	5.4	2471	2	T42977	large tegument pro
41	95	5.4	2253	2	T30336	nuclear/mitotic ap
42	94.5	5.4	1935	1	S06006	myosin beta heavy
43	94.5	5.4	1935	2	A59286	myosin heavy chain
44	94	5.3	1778	2	T50074	probable nucleopor
45	93.5	5.3	818	2	T31464	stage II sporulati

ALIGNMENTS

RESULT 1
T08667
hypothetical protein DKFZp547O0510.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C;Accession: T08667
R;Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16466
A;Accession: T08667
A;Molecule type: mRNA
A;Residues: 1-330 <POU>
A;Cross-references: EMBL:AL049929
A;Experimental source: fetal brain; clone DKFZp547O0510
C;Genetics:
A;Note: DKFZp547O0510.1

Query Match 94.6%; Score 1667; DB 2; Length 330;
Best Local Similarity 99.7%; Pred. No. 1.6e-114;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy      21 ILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVMV 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVMV 60

Qy      81 KGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVCKA 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVCKA 120

Qy     141 NSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDTSSLLSRH 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 NSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDTSSLLSRH 180

Qy     201 KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAV 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAV 240

Qy     261 VELVTVKSPDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVFNMVLWIMIALALA 320
           ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db     241 VELVTVKSPDTSLIRKTRTILEAKRAKNPASPYNLAYKYNFEYSVVFNMVLWIMIALALA 300

Qy     321 VITTSYNIWNNMDPGYDSIIYRMTNQKIRMD 350
           |||||||||||||||||||||||||||||||
Db     301 VITTSYNIWNNMDPGYDSIIYRMTNQKIRMD 330
```

RESULT 2
T23876
hypothetical protein RO3E1.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

```
C;Accession: T23876
R;McMurray, A.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19812
A;Accession: T23876
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-324 <WIL>
A;Cross-references: EMBL:Z92837; PIDN:CAB07401.1; GSPDB:GN00028; CESP:R03E1.2
A;Experimental source: clone R03E1
C;Genetics:
A;Gene: CESP:R03E1.2
A;Map position: X
A;Introns: 59/1; 180/2; 293/3

  Query Match             10.9%;  Score 192.5;  DB 2;  Length 324;
  Best Local Similarity   22.3%;  Pred. No. 9.6e-07;
  Matches   86;  Conservative   64;  Mismatches  134;  Indels  101;  Gaps   13;

Qy        4 FVVLL-ALVAGVLGNEFSILKSPGSVVFRNGNWPIP-------GERIPDVAALSM-GFSV 54
              ||:|: ||||     :    |:  :| |: | : :  :          | |   ::|   : ||||
Db        3 FVLLVSALVAACAASSLEIIAAPASLKFPDADATLKTTLISTLNENILGLSARQVDGFSV 62

Qy       55 KEDLSWPGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIH 114
                |          :| ||||  :| |  | |  | |    |       :        |       |
Db       63 SVD---------MFSRPRALAVVTVIGADNLNL-GGKTYSTKSDGVENVGFDQDMALIF 111

Qy      115 SLFSEETPV--------VLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLS 166
                |    |        :| ||  :| |              | :| ::  |: |
Db      112 GADRENVQVSAGGITGSMLALAAQQEAV-----------DATV-IKTKRSTL-------- 151

Qy      167 SLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRY 226
                                  : ||: :: ::: : :    :  |:: |:: : : ||  |
Db      152 -----------------VKELEAVYQLAAAI-KSAGVQMDNNADVFRVSITGLVGI---- 189

Qy      227 GEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELV--------------------T 265
                 :| |  :|   :  |::  ,   : : ||| ||:||:
Db      190 -TESAQKEEAVADVKAAIEALTASINNAYGGQAVIELLAFEAESGAAEKSTESREIPNHN 248

Qy      266 VKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVFNMVLWIMIALALAVIITS 325
              ::  ||     |:   |   |   | |                  :|   :| : | ::: | :|:|
Db      249 IQKRDTKSKTPTQLARENLQVTVPVSS---------DYPAIFAIFLGLVVILVVALIYIV 299

Qy      326 YNIWNMDPGYDSIIYRMTNQKIRMD 350
                 : ::||  ||||||||  ::: |
Db      300 VGMASIDPEKDSIIYRMTTTRMKKD 324

RESULT    3
T52336
probable type 1 membrane protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C;Accession: T52336
R;Goo, J.H.; Park, A.R.; Park, W.J.; Park, O.K.
Plant Mol. Biol. 41, 415-423, 1999
A;Title: Selection of Arabidopsis genes encoding secreted and plasma membrane proteins.
A;Reference number: Z25437; MUID:20064977
A;Accession: T52336
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-364 <GOO>
A;Cross-references: EMBL:AF104329; PIDN:AAD11797.1
C;Genetics:
A;Gene: PMP

  Query Match             6.9%;  Score 121.5;  DB 2;  Length 364;
  Best Local Similarity   22.0%;  Pred. No. 0.18;
  Matches   86;  Conservative   72;  Mismatches  154;  Indels   79;  Gaps   18;

Qy        3 VFVVLLALVAGVLGNEFSILKSPGSVVF-----------RNGNWPIPGERIPDVAALS-- 49
              |||| | |          |        :: ||| |              |:  :  :|    :    ||:|
Db        6 VFVVALLLTLNYRG------EASGSVFFIDGSNNQYLRPRSSSEALPMSPVEISAAVSAL 59

Qy       50 MGFSVKEDLSWPG------LAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVP 103
              :||:   |:  |          | | ||||  : : | : : |             |: | ||:
Db       60 LGFAPSATLTADGSSKLNKILKPNPFERPRAAFVLEIAGADDMLLETSPSHSF-LGNAIR 118

Qy      104 FSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRL----- 158
              |: |           :  :|    :|  :|  |  | : :|   |  :          : |
Db      119 SSIKS------DSYKADT----ELPDNEVVVVSVNEPSSDVTDKDI--NDFASWLGGSYV 166

Qy      159 --FQENSVLSSLPLNSLSRNNEVDL-------LFLSELQVLHDISSLLSRHKHLAK--DH 207
                 : :| | |:||      | | :|            |: | :  :|   :| :  |:   |
Db      167 AGAEPSSGLLSIPLAG-GANVEFNLEKEAERKFALNLLGLYQNIRQAVSVYDDLSHGIDR 225

Qy      208 SPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVK 267
              : :|        |:| : : ||:        :         :|:  | |  : : : : |    ::| |
Db      226 TAELTVGRFGGIDALAQEYGQG--MAKQGMDVLLSTLSKLFNLLETSHKG----QIVGVI 279

Qy      268 SFDTSLIRKTRTILEAKQAKNPA----------SPYNLAYKYNFEYSVVFNMVLWIM-IA 316
                | : ::: :|    ::: |           |    :|      |  :|    :  |:  |
Db      280 VLDERVNQESENLLNFGSSRSSARSMVEVEGIPSAAIIA-----EVILVRLTLAWLTGII 334

Qy      317 LALAVIITSYNIWNMDPGYDSIIYRMTNQKI 347
              | :| |:   |  : ||     |:::|  :| |:
Db      335 LLIATILGVYFLMNMPLTKDTLLY--SNVKL 363

RESULT    4
T40039
hypothetical protein SPAC9B6.13 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C;Accession: T40039; T40794
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21901
A;Accession: T40039
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-663 <WOO>
A;Cross-references: EMBL:AL031517; PIDN:CAA20649.1; GSPDB:GN00067; SPDB:SPBC28E12.02
A;Experimental source: strain 972h-; cosmid c28E12
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z21875
A;Accession: T40794
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-663 <WO2>
A;Cross-references: EMBL:AL049769; PIDN:CAB42374.1; GSPDB:GN00067; SPDB:SPAC9B6.13
A;Experimental source: strain 972h-; cosmid c9B6
C;Genetics:
A;Gene: SPBC28E12.02; SPAC9B6.13
A;Map position: 2

  Query Match             6.5%;  Score 114;  DB 2;  Length 663;
  Best Local Similarity   21.5%;  Pred. No. 1.5;
  Matches   62;  Conservative   43;  Mismatches   92;  Indels   92;  Gaps   13;

Qy       13 GVLGNEF-SILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHR 71
              ||  :|  ||: | ||:|  ||           ||:           | ||  : || |
Db      414 GVSSSELSSIVSSTGSIVETNG----IGEK--------MSFSPLKKLSIP---------- 451

Qy       72 P-----RATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQ 126
              |        :  :: :  |  ||  |  |           :  || ||  ::|  :     :     :     :|
Db      452 PTEFVAQIAIICMASGVEML-LKTNGIEYFGQENTVPIAMDKASKIFYKFGQSQWHQILL 510

Qy      127 LAPSEERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLS 185
               ||::::: :: ||  |    :  :       |    |:|                          |:||
Db      511 EAPTKDQDFISGKKNGKLDKVKQQCRFNLKNG---------------------DILFCP 548
```

```
Qy    186 ELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGL-------------------DEIGKR 225
          :   :  :                 |:|| ||  :                  :|| |:
Db    549 QSTSIFTV---------------DIYSDELERVIKGMNTMLLEFPAEMHFYVPEEIHKK 592

Qy    226 -YGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSL 273
            |  || :  :| |  :: :|:    || | ::   |   |  :|
Db    593 LIGFRGEQIQRVTK-LYNSYIEFSTTPFDCYGHNVLIR--TPSKFSENL 638
```

RESULT 5
T25243
hypothetical protein T24D5.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T25243
R;Wilkinson, J.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z20003
A;Accession: T25243
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-408 <WIL>
A;Cross-references: EMBL:Z68012; PIDN:CAA92020.1; GSPDB:GN00028; CESP:T24D5.2
A;Experimental source: clone T24D5
C;Genetics:
A;Gene: CESP:T24D5.2
A;Map position: X
A;Introns: 114/2; 134/3; 206/3; 309/3; 355/2
C;Superfamily: Caenorhabditis elegans hypothetical protein T24D5.2

```
  Query Match             6.2%;  Score 109.5;  DB 2;  Length 408;
  Best Local Similarity   20.8%;  Pred. No. 1.6;
  Matches   43;  Conservative   47;  Mismatches   84;  Indels   33;  Gaps    7;

Qy     81 KGVNK-LALPPGSVIS------YPLENAVPFSLDSVANSIHSL-FSEETPVVLQLAPSEE 132
           | :|: |  | | |::       : |:|   :  ::    : : |||    :|  :  :
Db    103 KAMNQFLKGPDGEVLAKYFQFCHELKNTTYKNKYTIRLDAYKMTFSERDDSILIRSDGDS 162

Qy    133 RVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHD 192
          |:|     |::| : :: : :||| |  |      | || : : |:          :|: :
Db    163 TVHMGISNNTLFRNRNMAIEELRNMLSQNIRVYKSLFVTNRSQQ-------MGEMDLSKC 215

Qy    193 ISSLLSRHKHLAKDHSPDLYSLELAGLDEIGK-------RYGE--DSEQFRDASKILVDA 243
          :|:: : :  :  |  |::  :           : |          ||    : || |:|: : :|
Db    216 VSTITAENLTVVMDSFPNIIEIVKLHRPMLNKFTLMCSGEYGNIMELEQVRNAANLKLDG 275

Qy    244 LQKFADDMYSLYGGNAVVELVTVKSFD 270
           | |  ||            :  :|  :|:
Db    276 LSKITDDQ---------LNSITATNFE 293
```

RESULT 6
S35732
dnaA protein - Spiroplasma citri
C;Species: Spiroplasma citri
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
C;Accession: S35732
R;Ye, F.; Laigret, F.; Bove, J.
submitted to the EMBL Data Library, December 1992
A;Description: Nucleotide sequence and genetic organization at the replication origin (c
A;Reference number: S35732
A;Accession: S35732
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-450 <YEF>
A;Cross-references: EMBL:Z19108; NID:g49345; PIDN:CAA79521.1; PID:g49346
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: replication initiation protein dnaA
C;Keywords: DNA binding; nucleotide binding; P-loop
F;146-153/Region: nucleotide-binding motif A (P-loop)
F;208-212/Region: nucleotide-binding motif B

```
  Query Match             6.0%;  Score 105;  DB 2;  Length 450;
  Best Local Similarity   20.4%;  Pred. No. 3.9;
  Matches   67;  Conservative   41;  Mismatches  115;  Indels  106;  Gaps   13;

Qy     81 KGVNKLALPPGSVIS---------------YPLENAVPFSLDSVANSIHSLFSEETPVV 124
           | :|| |     ||||                |    | || |       :| |       |
Db     83 KNINKQA----SVISKIDTLTENNNLAYYENYTFENFVR------GDSNHEAMQAALAVA 132

Qy    125 LQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENS--------VLSSLPLNSLSRN 176
           | |         ::: | :      |   :     | ::: |:              : :: |::
Db    133 LDLGKKWNPLFIYGDSGLGKTHLLHAIENKVNEIYKTNNRVKYLKADEFGKIAMDILNQG 192

Qy    177 NEV-----------DLLFLSELQVL-----------HDISSLLSRHKH--LAKDHSPDLY 212
          :|:             | | : ::|:|          | :| : ::|    :  |  ||
Db    193 HEIIEAFKTSYDIYDCLLIDDIQLLAKRNKTNELFFHIFNSYIEKNKQIVITSDKYPD-- 250

Qy    213 SLELAGLD-------EIGKRYGEDSEQFRDASKILVDALQ-----------------KF 247
            :| | :           |    | ||  |   | |||   |:                  |
Db    251 --DLGGFEARIISRFSYGLSIGLDSPDFETALKILEQKLKHQNNLGLFSEESLEFIALNF 308

Qy    248 ADDMYSLYGG--------------NAVVELVTV-KSFDTSLIRKTRTILEAKQAKNPASP 292
           |:  | |                | :: |  | |:|   : ::     |     |   :   |
Db    309 NSDVRKLEGAIKRLLFLAVMNKKPNEIITLADVEKAFKNAPLQNNEKITPKKIKQIVADS 368

Qy    293 YNLAYKYNFEYSVVFNMVLWIMIALALAV 321
           ||:  |      | | |     :| |  ||:
Db    369 YNTTIKAMMSKSRVSN----VMQARQLAM 393
```

RESULT 7
T19941
hypothetical protein C44H4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T19941
R;Smye, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19200
A;Accession: T19941
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-458 <WIL>
A;Cross-references: EMBL:Z79598; PIDN:CAB01867.1; GSPDB:GN00028; CESP:C44H4.1
A;Experimental source: clone C44H4
C;Genetics:
A;Gene: CESP:C44H4.1
A;Map position: X
A;Introns: 24/3; 157/1; 257/3; 372/3; 425/1

```
  Query Match             6.0%;  Score 105;  DB 2;  Length 458;
  Best Local Similarity   22.5%;  Pred. No. 4;
  Matches   69;  Conservative   47;  Mismatches  116;  Indels   74;  Gaps   12;

Qy     18 EFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVM 77
           :  ||  | :::               | : | | ||                    :|   : :
Db    149 QLKILSLPDNLI----------EYVQDNAFLS------------------YHSRDSLLK 179

Qy     78 VMVKGVNKLALPPGSVISYPLENAVPFSLD-SVANSIHSLFSEETPVVLQLAPSEERVYM 136
           : :   |  |: |  ::    |||     ||| :: : | |    |  | :   |:    |::
Db    180 LDLSANNLTAIHPTGLLG--LENLSQLSLDKNLLSEIPSQALENIPSLEDLSLGVNRIHT 237

Qy    137 VGKANSV----FEDLSVTLRQLR---NRLFQENSVLSSLPL--NSLSRNNEVDLLFLSEL 187
           : : ||:      : ||: : |:|    :  | |  :|| |  |  | |: :     |  :  |
Db    238 ISR-NSLPLPNLKSLSLEVNQIRLIPSDSFSETPLLSYLYLGNNLLTSIDASMFLHIGGL 296

Qy    188 QVL-----HDISSLLSRHK-----------HLAKDHSPDLYSLELAGLDEIGKRYGEDSE 231
           :||       ||:|: :  |               ||   |:| |   |||
Db    297 KVLSMSNNKDITSIQANGKLSVQQFKLCICILAFQHAPSLIRLELFDC------------ 344

Qy    232 QFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPAS 291
```

```
            |:|   :|||       |  | : ::  |  |  | :    | |:     :  :
Db    345 ---SISRIEPKSLQKVQHIQVILLSRNQITQINAVDDFAFSQL-PMLTSLDLSSNRLESL 400

Qy    292 PYNLAY 297
           | |: |
Db    401 PSNVIY 406
```

RESULT   8
G64685
hypothetical protein HP1327 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: G64685
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467
A;Accession: G64685
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-412 <TOM>
A;Cross-references: GB:AE000634; GB:AE000511; NID:g2314489; PIDN:AAD08377.1; PID:g231450

```
  Query Match             5.9%;  Score 104;  DB 2;  Length 412;
  Best Local Similarity   22.0%;  Pred. No. 4;
  Matches   70;  Conservative   55;  Mismatches  101;  Indels   92;  Gaps   15;

Qy     48 LSMGFSVKEDLSWPGLAVGNLFH-RPRATVMVMVKGVNKLALPPGSVISYPLEN------ 100
           :|:| | | ||:   |     :     :: : |   :||:   |::  :||
Db     97 MSLGLSQKVDLNGKKLTQSQMIDLEKQKKILELKKTKQQLAI---SLMINGIENYKNQQE 153

Qy    101 --AVPFSLDSVANSIHSLFSEETPVVLQLAPSE-ERVYMVGKANSVFEDLSVTLRQLRNR 157
              :  :: :: |:::      :| :: || |   :   :  | |:: | || :    :
Db    154 IELLKTAIKNLENTLYQANHSSSPNLTAIAKLEILKSQLEIKKNNLEEALSASHYSMGEL 213

Qy    158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELA 217
           |:|| :|| |  |    | |          | ||:||:              | : :|
Db    214 AFKENELLSIAPKN-FEFNRE---------QELHNISA-------------TNYDIAIA 249

Qy    218 GLDEIGKRYGEDSEQFRDASK--ILVD---------ALQKFADDMYS--------LYGG 257
            |||       | |::    :|   | |              : | :  ||:|       :||
Db    250 RLDE------EKSQKDITLAKKSFLEDVNVTGVYYFRSKQYYNYDMFSIALSIPLPIYGK 303

Qy    258 NAVVELVTVKSFDT------------------SLIRKTRTI---LEA-----KQAKNPA 290
            |  :|| |   ::                   |::|  |:    ||:        || :  |
Db    304 QA--KLVEQKKKESLVFKSEVENTKNKTHHLALKLLKKLETLQKNLESINKIIKQNEKIA 361

Qy    291 SPYNLAYKYNFEYSVVFN 308
             | |  | | :|:  :|
Db    362 QIYALDLKSNGDYNAYYN 379
```

RESULT   9
F64543
conserved hypothetical secreted protein HP0190 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: F64543
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467
A;Accession: F64543
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-502 <TOM>
A;Cross-references: GB:AE000539; GB:AE000511; NID:g2313275; PIDN:AAD07257.1; PID:g231
C;Genetics:
A;Start codon: TTG

```
  Query Match             5.9%;  Score 103.5;  DB 2;  Length 502;
  Best Local Similarity   22.9%;  Pred. No. 5.8;
  Matches   72;  Conservative   35;  Mismatches  115;  Indels   93;  Gaps   15;

Qy     93 VISYPLENA------VPFSLDSVANSIHSLFSEETPVVLQLAPSEERV--------YMVG 138
           ||:  | ||        |    ||   | : | ||:       : |     : :         |:
Db     95 VIAKELLNAANRGVKVRILLDD--NGLDSDFSD-----IMLLNFHKNIEVKIFNPYYIRN 147

Qy    139 KANSVFE---DLSVTLRQLRNRLFQENSVLSSLPLNSLSRN---NEVDLLFLSELQVL-- 190
           |     ||   |       ::: |:|| ::    :    ::  |    |::|  || :|  |
Db    148 KGLRYFEMLADYERIKKRMHNKLFIVDNFAVIIGGRNIGDNYFDNDLDTNFL-DLDALFF 206

Qy    191 ------------------HDIS-SLLSFHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSE 231
                             |  |||  || |   :         : |:| | |      ||
Db    207 GGVASKAKESFERYWRFHRSIPVSLLRTHKRLKNN------AKEIAKLHEKIPISAEDKN 260

Qy    232 QFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPAS 291
           ||        :|   ||:     | :| |||:      :  |  |||  |       :   | |  ||
Db    261 QFEKKVNDFIDRFQKY---QYPIYYGNAIFLADSPKKIDTPLYSPIKIAFE-KALKNAKD 316

Qy    292 PYNLAYKY-----------------NFEYSVVFNMVLWIMIALALAVIITSYNIWNMDPG 334
             :| |                     | ::: |          :|:      |  |  |
Db    317 SVFIASSYFIPGKKMMKIFKNQISKGIELNILTN-------SLSSTDAIVVYGAWE---- 365

Qy    335 YDSIIYRMTNQKIRM 349
                |  || :||
Db    366 ------RYRNQLVRM 374
```

RESULT   10
T43619
secreted protein kinase ypkA - Yersinia pestis plasmid pCD1
N;Alternate names: protein kinase A homolog
C;Species: Yersinia pestis
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T43619; T42850
R;Hu, P.; Elliott, J.; McCready, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Bruba
J. Bacteriol. 180, 5192-5202, 1998
A;Title: Structural organization of virulence-associated plasmids of Yersinia pestis.
A;Reference number: Z22578; MUID:98422474
A;Accession: T43619
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-732 <HUP>
A;Cross-references: EMBL:AF053946; NID:g2996222; PIDN:AAC62602.1; PID:g2996279
A;Experimental source: strain KIM
R;Perry, R.D.; Straley, S.C.; Fetherston, J.D.; Rose, D.J.; Gregor, J.; Blattner, F.R
Infect. Immun. 66, 4611-4623, 1998
A;Title: DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of Yersini
A;Reference number: Z22273; MUID:98427122
A;Accession: T42850
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-732 <PER>
A;Cross-references: EMBL:AF074612; NID:g3822037; PIDN:AAC69765.1; PID:g3822045
A;Experimental source: strain KIM5
C;Genetics:
A;Gene: ypkA
A;Genome: plasmid pCD1

```
  Query Match             5.9%;  Score 103.5;  DB 2;  Length 732;
  Best Local Similarity   21.2%;  Pred. No. 10;
  Matches   73;  Conservative   51;  Mismatches  103;  Indels  117;  Gaps   17;

Qy     12 AGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSM---------GFSVKEDLSWPG 62
           |||: |:       ||:|||    :     ||   |  |  | :             ||:   |    |
```

```
Db    264 AGVVHNDI----KPGNVVFDRAS----GE--PVVIDLGLHSRSGEQPKGFT--ESFKAPE 311

Qy     63 LAVGNLFHRPRATVMVMV------------------KGVNKLALPPGSVISYPLENAVP 103
          | ||||     :: | ::|                  :|:  :   |   |:    ||  |
Db    312 LGVGNLGASEKSDVFLVVSTLLHCIEGFEKNPEIKPNQGLRFITSEPAHVMD---ENGYP 368

Qy    104 FSLDSVANSIHSLFSEETPVVLQLA------PSEERVYMVGKANSVFEDLSVTLRQLRNR 157
                :|  : : ::      :| ::        :| |::      :: |: :  :  |::
Db    369 IHRPGIA-GVETAYTRFITDILGVSADSRPDSNEARLHEFLSDGTIDEESAKQI--LKDT 425

Qy    158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL------ 211
           |  | | ||:              |:  ::  : | ::| ||  |   |     |:
Db    426 LTGEMSPLST------------DVRRITP-KKLRELSDLLRTHLSSAATKQLDMGGVLSD 472

Qy    212 YSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDT 271
                |  ||:  :   |  | :|                              :|||:
Db    473 LDTMLVALDKAEREGGVDKDQ-------------------------------LKSFN- 498

Qy    272 SLIRKTRTILE---------AKQAKNPASPYNLAYKYNFEYSVV 306
          ||| ||   ::|           | :    |||    :: ||   |:|
Db    499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539
```

RESULT 11
S30060
protein kinase ypkA - Yersinia pseudotuberculosis
C;Species: Yersinia pseudotuberculosis
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C;Accession: S30060
R;Galyov, E.E.; Hakansson, S.; Forsberg, A.; Wolf-Watz, H.
Nature 361, 730-732, 1993
A;Title: A secreted protein kinase of Yersinia pseudotuberculosis is an indispensable vi
A;Reference number: S30060; MUID:93180911
A;Accession: S30060
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-732 <GAL>
A;Cross-references: EMBL:X69439; NID:g49186; PIDN:CAA49215.1; PID:g49187
C;Keywords: serine/threonine-specific protein kinase

```
  Query Match             5.9%;  Score 103.5;  DB 2;  Length 732;
  Best Local Similarity   21.2%;  Pred. No. 10;
  Matches   73;  Conservative   51;  Mismatches  103;  Indels  117;  Gaps   17;

Qy     12 AGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSM---------GFSVKEDLSWPG 62
          |||: |:        ||:|||  :      ||   |  |  |  |  :      ||:  |     |
Db    264 AGVVHNDI----KPGNVVFDRAS----GE--PVVIDLGLHSRSGEQPKGFT--ESFKAPE 311

Qy     63 LAVGNLFHRPRATVMVMV------------------KGVNKLALPPGSVISYPLENAVP 103
          | ||||     :: | ::|                  :|:  :   |   |:    ||  |
Db    312 LGVGNLGASEKSDVFLVVSTLLHCIEGFEKNPEIKPNQGLRFITSEPAHVMD---ENGYP 368

Qy    104 FSLDSVANSIHSLFSEETPVVLQLA------PSEERVYMVGKANSVFEDLSVTLRQLRNR 157
                :|  : : ::      :| ::        :| |::      :: |: :  :  |::
Db    369 IHRPGIA-GVETAYTRFITDILGVSADSRPDSNEARLHEFLSDGTIDEESAKQI--LKDT 425

Qy    158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL------ 211
           |  | | ||:              |:  ::  : | ::| ||  |   |     |:
Db    426 LTGEMSPLST------------DVRRITP-KKLRELSDLLRTHLSSAATKQLDMGGVLSD 472

Qy    212 YSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDT 271
                |  ||:  :   |  | :|                              :|||:
Db    473 LDTMLVALDKAEREGGVDKDQ-------------------------------LKSFN- 498

Qy    272 SLIRKTRTILE---------AKQAKNPASPYNLAYKYNFEYSVV 306
          ||| ||   ::|           | :    |||    :: ||   |:|
Db    499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539
```

RESULT 12
C84538
probable LRR receptor protein kinase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84538
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Fallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: C84538
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-925 <STO>
A;Cross-references: GB:AE002093; NID:g4544402; PIDN:AAD22312.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g16250
A;Map position: 2

```
  Query Match             5.9%;  Score 103.5;  DB 2;  Length 925;
  Best Local Similarity   22.6%;  Pred. No. 14;
  Matches   86;  Conservative   43;  Mismatches  126;  Indels  125;  Gaps   16;

Qy      4 FVVLL--------ALVAGVLGNEFS----ILKSPGSVVFRNGNWPIPGERIPD------- 44
          ||:||          :| |   : ||    :|    |:  |  :||| |:    |
Db     11 FVLLLLCLVLFFDCVVVGQTQSRFSEKLILLNLRSSLGLRGTDWPIKGDPCVDWRGIQCE 70

Qy     45 ----------------VAALSMGFSV-------------KEDLSWPGLAVGNLFHRPRAT 75
                          :  |:  |||                    |: ||  :    |
Db     71 NGSIIGINISGFRRTRIGKLNPQFSVDPLRNLTRLSYFNASGLALPG-TIPEWF------ 123

Qy     76 VMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVA---------NSIHSLFSEETPVVLQ 126
                 ||: |||      : |   :      |||:| ::                ||: ||          :|
Db    124 ------GVSLLALEVLDLSSCSVNGVVPFTLGNLTSLRTLNLSQNSLTSLVPSSLGQLLN 177

Qy    127 LAPSE-ERVYMVGKANSVFEDLS--VTLRQLRNRL-----------------FQENSVL 165
          |:  :    |    |      |   |    :||       ||                        |  ||
Db    178 LSQLDLSRNSFTGVLPQSFSSLKNLLTLDVSSNYLTGPIPPGLGALSKLIHLNFSSNSFS 237

Qy    166 SSLP--------------LNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPD 210
           | :|                 :|||  :    :|   ||:||::         :|||      :        |
Db    238 SPIPSELGDLVNLVDFDLSINSLSGSVPQELRKLSKLQLMAIGDNLLSGTLPV------D 291

Qy    211 LYSLELAGLDEIGKRYGEDSEQFRDAS------KILVDALQKFADDM-YSLYGGNAVVEL 263
          |:|  |  :  |   :   |      |       :||    |     |    : ||  |   : : |:
Db    292 LFSAE-SQLQTLVLRENGFSGSLPDVCWSLPKLRILDIAKNNFTGLLPYSSYDSDQIAEM 350

Qy    264 VTVKSFD-----TSLIRKTR 278
           | : |           | ::|: |
Db    351 VDISSNTFYGELTPILRRFR 370
```

RESULT 13
A71685
guanosine pentaphosphate phosphohydrolase (gppA) RP294 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: A71685
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: A71685
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-470 <AND>
A;Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14755.1; PID:g386
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: gppA; RP294

```
  Query Match             5.8%;  Score 103;  DB 2;  Length 470;
```

```
  Best Local Similarity   21.7%;  Pred. No. 5.8;
  Matches   65;  Conservative   56;  Mismatches  123;  Indels   56;  Gaps   12;

Qy     45 VAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVMVKGVNKL-ALPPGSVISYPLENAVP 103
          :: :|  | :  ||   | : ::::          | | || :|| |: |
Db    117 ISGISDAFGIVADLGGGSLELAHIYN---------KKVGKLKSLPLGTKI--------- 157

Qy    104 FSLDSVANS-------IHSLFSEETPVVLQLAPSEERVYMVGKA-----NSVFEDLSVTL 151
               :|||       | :  || |        :|::| |        | :: |
Db    158 -----IANSNFSDVGLITKMLEEEFGVA-----HYPNLYLIGGALRLMSRIYMESINYPL 207

Qy    152 RQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL 211
          : | |  |: | |  | |  ||: ::: | :  :  : ::  ::  | : |  ||:
Db    208 KNLHN--FEINRVEFELYLEKLSQIDKLKLSYYEQKAINYNAVLVI---KAMIKVFSPEK 262

Qy    212 YSLELAGLDEIGKRYGEDSEQFRDASKIL----VDALQKFADDMYSLYGGNAVVELVTVK 267
             :  | | | |: ||  : :  | :      |  |  | :: :        :: : :
Db    263 IIISNYGLKE-GVRF--DSLPYHETEKDIIYERVKRLVNFDRNICKIEKYIEALQYLLIN 319

Qy    268 SFDTSLIRKTRTILEAKQAKN--PASPYNLAYKYNFEYSVVFNMVLWIMIALALAVIITS 325
           |  |:||     |: |:  ||        |  ::      : |:      :|: :||| |  |:
Db    320 SDATTLIIIELAIMLAQYNKNIDKTLRANFVSEFILSSDIPFSHRQRLMLGIALTVTYTA 379
```

RESULT 14
JQ0894
P115 protein - Mycoplasma hyorhinis
C;Species: Mycoplasma hyorhinis
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 02-Feb-2001
C;Accession: JQ0894
R;Notarnicola, S.M.; McIntosh, M.A.; Wise, K.S.
Gene 97, 77-85, 1991
A;Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding
A;Reference number: JQ0894; MUID:91138990
A;Accession: JQ0894
A;Molecule type: DNA
A;Residues: 1-979 <NOT>
A;Cross-references: GB:M34956
A;Note: the authors translated the codon AAA for residue 956 as Leu
C;Comment: This protein is located in the cytoplasm.
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: chromosome segregation protein SMC1
C;Keywords: nucleotide binding; P-loop
F;32-39/Region: nucleotide-binding motif A (P-loop)

```
  Query Match             5.8%;  Score 103;  DB 2;  Length 979;
  Best Local Similarity   19.6%;  Pred. No. 17;
  Matches   67;  Conservative   63;  Mismatches  126;  Indels   86;  Gaps   15;

Qy     11 VAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKE------------D 57
          :| :| ::::|:   |  :|| |     | ::   : |:    :||             |
Db    537 IAKILDHKYTIVSLEGD-LFRPGGTITGGSKLERTSILNYDIKIKEHTNTLKFAEDQIHD 595

Qy     58 LSWPGLAVGNLFHRPRATVM-------------VMVKGVNKLALPPGSVISYPLEN--A 101
          |      : |      :|:              :: : :| | |         |:  :
Db    596 LKIKQQTIYNEIETVNSTIQQVKIEANSINSKLNILNEELNNLKLNASEIFKEQQEDQES 655

Qy    102 VPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMV----GKANSVFEDLSVTLRQLRNR 157
          :  | ||    :|    |  |    ::|    ::|: :      ||  :  ::|    ||:| :
Db    656 LNLSFDSEKLNIEKQISTLT---IELNSKKDRLTNLISEQGKGETKKQELDAKLRKLNTQ 712

Qy    158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELA 217
                     :|::  |    |         :          || |  |  : : : |||:
Db    713 ----------HSDSITEQNRAKFL-------VEQNQKRLSEHYKLTLEAASEQYSLD-- 752

Qy    218 GLDEIGKRYGEDS--EQFRDASKILVDALQKFADDMYSLYGGNA--VVELVTVKS----- 268
           ||   |:   ||   :: ::    : ::|: :|  :::     |          :  || |  ||
Db    753 -LDIEQARHFVDSLKKELKELGNVNLEAITEF-EEVNQRYQEKKQYIEELTTAKSKIEEA 810

Qy    269 ---FDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVF 307
             |  :| || |:              ||     | |:::||

Db    811 ISDLDKIIINKTTEIV-----------NLV---NNEFNMVF 837
```

RESULT 15
A39611
probable GTP-binding protein - mouse
N;Alternate names: gb110/Mov protein
C;Species: Mus musculus (house mouse)
C;Date: 13-Sep-1991 #sequence_revision 13-Sep-1991 #text_change 02-Feb-2001
C;Accession: A39611; I48358; S38628
R;Mooslehner, K.; Mueller, U.; Karls, U.; Hamann, L.; Harbers, K.
Mol. Cell. Biol. 11, 886-893, 1991
A;Title: Structure and expression of a gene encoding a putative GTP-binding protein i
A;Reference number: A39611; MUID:91117255
A;Accession: A39611
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1004 <MOO>
A;Cross-references: EMBL:X52574; NID:g53168; PIDN:CAA36803.1; PID:g53169
R;Hamann, L.; Bayer, K.U.; Jensen, K.; Harbers, K.
Mol. Cell. Biol. 14, 5786-5793, 1994
A;Title: Interaction of several related GC-box- and GT-box-binding proteins with the
A;Reference number: I48358; MUID:94344088
A;Accession: I48358
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-45 <RES>
A;Cross-references: EMBL:X75819; NID:g415391; PIDN:CAA53453.1; PID:g433685
C;Genetics:
A;Gene: gb110
C;Keywords: nucleotide binding; P-loop
F;525-532/Region: nucleotide-binding motif A (P-loop)

```
  Query Match             5.8%;  Score 102.5;  DB 2;  Length 1004;
  Best Local Similarity   27.9%;  Pred. No. 19;
  Matches   56;  Conservative   19;  Mismatches   67;  Indels   59;  Gaps    9;

Qy    149 VTLRQLRNRLFQENSVLSS---------LPLNSLSRNNEVD---LLFLSELQVLHDISS 195
           : ||||  | |  |: ::           |     ||| ||    || | |||: |||
Db    299 ILLRQLLPTLLQGPSIFTAPKEVAEIKAQLETTLKSRNYEVKLRLLLHLEELQMEHDI-- 356

Qy    196 LLSRHKHLAK--------DHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKF 247
              || |            | :| | :||: |: |         |       :|      |:
Db    357 ---RHYDLDSVPMTWDPVDQNPRLLTLEVPGVAESRPSVLRGDHLF----ALLSSETQQ- 408

Qy    248 ADDMYSLYGGNAVVELVTVK-SFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYS-- 304
           ||   :   |      |||  || || |||: :                   | :| || ::
Db    409 -DDPVTYKGFVHKVELDRVKLSFSTSLLSRFVD-------------GLTFKVNFTFNRQ 453

Qy    305 ---------VVFNMVLWIMI 315
                   :     ||| |:
Db    454 PLRVQHRALELTGRWVLWPML 474
```

Search completed: July 3, 2001, 16:19:18
Job time: 155 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:17:23 ; Search time 12.1 Seconds
(without alignments)
990.861 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVPVVLLALVAGVLGNEFS.....MDPGYDSIIYRMTNQKIRMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	513	29.1	100	1	VATN_HUMAN	O75787 homo sapien
2	122	6.9	30	1	VATN_BOVIN	P81134 bos taurus
3	105	6.0	450	1	DNAA_SPICI	P34028 spiroplasma
4	103.5	5.9	502	1	Y190_HELPY	P56117 helicobacte
5	103.5	5.9	732	1	YPKA_YERPS	Q05608 yersinia ps
6	103	5.8	979	1	P115_MYCHR	P41508 mycoplasma
7	102.5	5.8	1004	1	MV10_MOUSE	P23249 mus musculu
8	100.5	5.7	1358	1	SIR4_YEAST	P11978 saccharomyc
9	100.5	5.7	2839	1	NF1_HUMAN	P21359 homo sapien
10	99	5.6	483	1	SYD_MYCCA	Q48979 mycoplasma
11	98.5	5.6	1935	1	MYSB_HUMAN	P12883 homo sapien
12	97.5	5.5	884	1	SECA_PORPU	P51381 porphyra pu
13	97.5	5.5	2841	1	NF1_MOUSE	Q04690 mus musculu
14	97	5.5	715	1	ORC3_MOUSE	Q91k30 mus musculu
15	96.5	5.5	805	1	SUS1_TULGE	Q41608 tulipa gesn
16	96.5	5.5	1934	1	MYSB_MESAU	P13540 mesocricetu
17	96.5	5.5	1938	1	MYSA_RAT	P02561 rattus norv
18	95.5	5.4	716	1	MYSB_RABIT	P04461 oryctolagus
19	94.5	5.4	1935	1	MYSB_PIG	P79293 sus scrofa
20	94.5	5.4	1935	1	MYSB_RAT	P02564 rattus norv
21	93.5	5.3	1939	1	MYSA_HUMAN	P11533 homo sapien
22	93	5.3	640	1	DXS_SYNY3	P73067 synechocyst
23	92.5	5.2	502	1	Y190_HELPJ	Q9zmp2 helicobacte
24	91.5	5.2	896	1	BOSS_DROME	P22815 drosophila
25	91.5	5.2	4092	1	DYHC_YEAST	P36022 saccharomyc
26	91	5.2	262	1	CUT8_SCHPO	P38937 schizosacch
27	91	5.2	427	1	FOLD_YEAST	Q12676 saccharomyc
28	90.5	5.1	465	1	MYSA_RABIT	P04460 oryctolagus
29	90.5	5.1	472	1	ATPB_RHOCA	P72247 rhodobacter
30	90.5	5.1	1102	1	MYSC_CHICK	P29616 gallus gall
31	90.5	5.1	1938	1	MYSA_MOUSE	Q02566 mus musculu
32	90	5.1	507	1	RECN_CAMJE	Q9phm5 campylobact
33	90	5.1	626	1	CEIB_ECOLI	P04479 escherichia
34	90	5.1	1032	1	KINN_HUMAN	Q12840 homo sapien
35	90	5.1	1085	1	CUT7_SCHPO	P24339 schizosacch
36	89.5	5.1	642	1	DEAD_KLEPN	P33906 klebsiella
37	89.5	5.1	968	1	Y682_METJA	Q58095 methanococc
38	89.5	5.1	3106	1	LMA2_MOUSE	Q60675 mus musculu
39	89	5.0	500	1	TACY_CLOPE	P19995 clostridium
40	89	5.0	904	1	PMS1_YEAST	P14242 saccharomyc
41	88.5	5.0	1663	1	CO3_HUMAN	P01024 homo sapien
42	88	5.0	442	1	THDF_MYCGE	P47254 mycoplasma
43	88	5.0	618	1	VE1_HPV63	Q07847 human papil
44	88	5.0	1177	1	MFD_BACSU	P37474 bacillus su
45	87.5	5.0	505	1	HAS1_YEAST	Q03532 saccharomyc

ALIGNMENTS

RESULT 1
VATN_HUMAN
ID VATN_HUMAN STANDARD; PRT; 100 AA.
AC Q75787;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE MEMBRANE SECTOR ASSOCIATED PROTEIN M8-9
DE (V-ATPASE M8.9 SUBUNIT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ludwig J., Schagger H., Boyd A., Apps D.K.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; Y17975; CAA76984.1; -.
KW Hydrolase; Hydrogen ion transport; Transmembrane.
FT TRANSMEM 53 73 POTENTIAL.
SQ SEQUENCE 100 AA; 11575 MW; A0705DB7E665P6DC CRC64;

Query Match 29.1%; Score 513; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.6e-31;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    251 MYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVFNMV 310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVFNMV 60

Qy    311 LWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
          |||||||||||||||||||||||||||||||||||||||||
Db     61 LWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 100
```

RESULT 2
VATN_BOVIN
ID VATN_BOVIN STANDARD; PRT; 30 AA.
AC P81134;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```
DE   VACUOLAR ATP SYNTHASE MEMBRANE SECTOR ASSOCIATED PROTEIN M8-9
DE   (V-ATPASE M8.9 SUBUNIT) (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Adrenal medulla;
RX   MEDLINE=98225166; PubMed=9556572;
RA   Ludwig J., Kerscher S., Brandt U., Pfeiffer K., Getlawi F., Apps D.K.,
RA   Schagger H.;
RT   "Identification and characterization of a novel 9.2-kDa membrane
RT   sector-associated protein of vacuolar proton-ATPase from chromaffin
RT   granules.";
RL   J. Biol. Chem. 273:10939-10947(1998).
CC   -!- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC       OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC   -!- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
KW   Hydrolase; Hydrogen ion transport; Transmembrane.
FT   NON_TER       1      1
FT   TRANSMEM     22    >30       POTENTIAL.
FT   NON_TER      30     30
SQ   SEQUENCE   30 AA;  3585 MW;  EA9FB26DCFA7D665 CRC64;

  Query Match             6.9%;  Score 122;  DB 1;  Length 30;
  Best Local Similarity   73.3%;  Pred. No. 0.0018;
  Matches   22;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     282 EAKQAKNPASPYNLAYKYNFEYSVVFNMVL 311
             | ||  |:|:: ||||||||||| ||||:||
Db       1 ETKQVKDPSTTYNLAYKYNFEYPVVFNLVL 30

RESULT   3
DNAA_SPICI
ID   DNAA_SPICI     STANDARD;      PRT;   450 AA.
AC   P34028;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA.
GN   DNAA.
OS   Spiroplasma citri.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Spiroplasmataceae; Spiroplasma.
OX   NCBI_TaxID=2133;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27556 / R8A2;
RX   MEDLINE=94290261; PubMed=7764984;
RA   Ye F., Renaudin J., Bove J.M., Laigret F.;
RT   "Cloning and sequencing of the replication origin (oriC) of the
RT   Spiroplasma citri chromosome and construction of autonomously
RT   replicating artificial plasmids.";
RL   Curr. Microbiol. 29:23-29(1994).
RN   [2]
RP   SEQUENCE OF 147-411 FROM N.A.
RC   STRAIN=ATCC 27556 / R8A2;
RX   MEDLINE=94109679; PubMed=8282191;
RA   Suzuki K., Miyata M., Fukumura T.;
RT   "Comparison of the conserved region in the dnaA gene from three
RT   mollicute species.";
RL   FEMS Microbiol. Lett. 114:229-233(1993).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION
CC       OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
CC       IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
CC       (DNAA BOX): 5'-TTATC(C/A)A(C/A)A-3'. DNAA BINDS TO ATP AND TO
CC       ACIDIC PHOSPHOLIPIDS.
CC   -!- SIMILARITY: BELONGS TO THE DNAA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z19108; CAA79521.1; -.
DR   EMBL; D14985; BAA03630.1; -.
DR   PIR; S35732; S35732.
DR   InterPro; IPR001957; -.
DR   Pfam; PF00308; bac_dnaA; 1.
DR   PRINTS; PR00051; DNAA.
DR   PROSITE; PS01008; DNAA; 1.
KW   DNA replication; DNA-binding; ATP-binding.
FT   NP_BIND     146    153       ATP (PROBABLE).
SQ   SEQUENCE   450 AA;  51537 MW;  C401EF9B157FB961 CRC64;

  Query Match             6.0%;  Score 105;  DB 1;  Length 450;
  Best Local Similarity   20.4%;  Pred. No. 1.6;
  Matches   67;  Conservative   41;  Mismatches  115;  Indels  106;  Gaps   13;

Qy      81 KGVNKLALPPGSVIS---------------YPLENAVPFSLDSVANSIHSLFSEETPVV 124
Db      83 KNINKQA----SVISKIDTLTENNNLAYYENYTFENFVR------GDSNHEAMQAALAVA 132

Qy     125 LQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENS--------VLSSLPLNSLSRN 176
Db     133 LDLGKKWNPLFIYGDSGLGKTHLLHAIENKVNEIYKTNNRVKYLKADEFGKIAMDILNQG 192

Qy     177 NEV-----------DLLFLSELQVL-----------HDISSLLSRHKH--LAKDHSPDLY 212
Db     193 HEIIEAFKTSYDIYDCLLIDDIQLLAKRNKTNELFFHIFNSYIEKNKQIVITSDKYPD-- 250

Qy     213 SLELAGLD-------EIGKRYGEDSEQFRDASKILVDALQ-----------------KF 247
Db     251 --DLGGFEARIISRFSYGLSIGLDSPDFETALKILEQKLKHQNNLGLFSEESLEFIALNF 308

Qy     248 ADDMYSLYGG-------------NAVVELVTV-KSFDTSLIRKTRTILEAKQAKNPASP 292
Db     309 NSDVRKLEGAIKRLLFLAVMNKKPNEIITLADVEKAFKNAPLQNNEKITPKKIKQIVADS 368

Qy     293 YNLAYKYNFEYSVVFNMVLWIMIALALAV 321
Db     369 YNITIKAMMSKSRVSN----VMQARQLAM 393

RESULT   4
Y190_HELPY
ID   Y190_HELPY     STANDARD;      PRT;   502 AA.
AC   P56117;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN HP0190.
GN   HP0190.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
```

```
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPDASE D FAMILY. CARDIOLIPIN
CC       SYNTHASE SUBFAMILY. STRONG, TO E.COLI YMDC.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000539; AAD07257.1; -.
DR   TIGR; HP0190; -.
DR   InterPro; IPR001736; -.
DR   Pfam; PF00614; PLDc; 2.
KW   Hypothetical protein; Transferase; Transmembrane.
FT   TRANSMEM      1     21       POTENTIAL.
SQ   SEQUENCE   502 AA;  58287 MW;  C80911F2FB59C714 CRC64;


  Query Match              5.9%;  Score 103.5;  DB 1;  Length 502;
  Best Local Similarity   22.9%;  Pred. No. 2.5;
  Matches   72;  Conservative   35;  Mismatches  115;  Indels   93;  Gaps   15;

Qy       93 VISYPLENA------VPFSLDSVANSIHSLFSEETPVVLQLAPSEERV--------YMVG 138
               ||:  | ||        |   ||    | : | ||:     : |     : :        |:
Db       95 VIAKELLNAANRGVKVRILLDD--NGLDSDFSD-----IMLLNFHKNIEVKIFNPYYIRN 147

Qy      139 KANSVFE---DLSVTLRQLRNRLFQENSVLSSLPLNSLSRN---NEVDLLFLSELQVL-- 190
               |    ||   |      ::: |:||   ::      :    ::   |     |::|   || :|  |
Db      148 KGLRYFEMLADYERIKKRMHNKLFIVDNFAVIIGGRNIGDNYFDNDLDTNFL-DLDALFF 206

Qy      191 -----------------HDIS-SLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSE 231
                                 |  |||   ||  |   :         : |:| |  |        ||
Db      207 GGVASKAKESFERYWRFHRSIPVSLLRTHKRLKNN------AKEIAKLHEKIPISAEDKN 260

Qy      232 QFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPAS 291
               ||         :|   ||:    | :| |||:     : |   | ||  |     :     |  |   ||
Db      261 QFEKKVNDFIDRFQKY---QYPIYYGNAIFLADSPKKIDTPLYSPIKIAFE-KALKNAKD 316

Qy      292 PYNLAYKY----------------NFEYSVVFNMVLWIMIALALAVIITSYNIWNMDPG 334
                 :|  |                  | ::: |        :|:        |    |    |
Db      317 SVFIASSYFIPGKKMMKIFKNQISKGIELNILTN-------SLSSTDAIVVYGAWE---- 365

Qy      335 YDSIIYRMTNQKIRM 349
                       |  ||  :||
Db      366 ------RYRNQLVRM 374


RESULT   5
YPKA_YERPS
ID   YPKA_YERPS     STANDARD;      PRT;   732 AA.
AC   Q05608;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROTEIN KINASE YPKA PRECURSOR (EC 2.7.1.-).
GN   YPKA.
OS   Yersinia pseudotuberculosis.
OG   Plasmid pIB1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=633;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YPIII;
RX   MEDLINE=93180911; PubMed=8441468;
RA   Galyov E.E., Haakansson S., Forsberg A., Wolf-Watz H.;
RT   "A secreted protein kinase of Yersinia pseudotuberculosis is an
RT   indispensable virulence determinant.";
RL   Nature 361:730-732(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YPIII;
RX   MEDLINE=94321324; PubMed=8045884;
RA   Galyov E.E., Haakansson S., Wolf-Watz H.;
RT   "Characterization of the operon encoding the YpkA Ser/Thr protein
RT   kinase and the YopJ protein of Yersinia pseudotuberculosis.";
RL   J. Bacteriol. 176:4543-4548(1994).
CC   -!- FUNCTION: ACTS AS A VIRULENCE DETERMINANT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X69439; CAA49215.1; -.
DR   EMBL; L33833; AAA68487.1; -.
DR   PIR; S30060; S30060.
DR   InterPro; IPR000719; -.
DR   InterPro; IPR002290; -.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Virulence; Plasmid; Transferase; Serine/threonine-protein kinase;
KW   ATP-binding; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    732       PROTEIN KINASE YPKA.
FT   DOMAIN      136    408       PROTEIN KINASE.
FT   ACT_SITE    270    270       BY SIMILARITY.
SQ   SEQUENCE   732 AA;  81707 MW;  7C09BDD4F9C9CBDF CRC64;


  Query Match              5.9%;  Score 103.5;  DB 1;  Length 732;
  Best Local Similarity   21.2%;  Pred. No. 4.2;
  Matches   73;  Conservative   51;  Mismatches  103;  Indels  117;  Gaps   17;

Qy       12 AGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSM---------GFSVKEDLSWPG 62
               |||: |:       ||:|||   :     ||  |  |   | :            ||:  |     |
Db      264 AGVVHNDI----KPGNVVFDRAS----GE--PVVIDLGLHSRSGEQPKGFT--ESFKAPE 311

Qy       63 LAVGNLFHRPRATVMVMV------------------KGVNKLALPPGSVISYPLENAVP 103
               | ||||      :: | ::|                   :|:  :     |  |:    || |
Db      312 LGVGNLGASEKSDVFLVVSTLLHCIEGFEKNPEIKPNQGLRFITSEPAHVMD---ENGYP 368

Qy      104 FSLDSVANSIHSLFSEETPVVLQLA------PSEERVYMVGKANSVFEDLSVTLRQLRNR 157
                  :|  : : ::       :| ::         :| |::         :: |: :   :   |::
Db      369 IHRPGIA-GVETAYTRFITDILGVSADSRPDSNEARLHEFLSDGTIDEESAKQI--LKDT 425

Qy      158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL------ 211
               |  | | ||:                 |:  ::   : | ::| ||   |     |        |:
Db      426 LTGEMSPLST-----------DVRRITP-KKLRELSDLLRTHLSSAATKQLDMGGVLSD 472

Qy      212 YSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDT 271
                   |  ||:  :  | |  :|                                     :|||:
Db      473 LDTMLVALDKAEREGGVDKDQ-------------------------------LKSFN- 498

Qy      272 SLIRKTRTILE---------AKQAKNPASPYNLAYKYNFEYSVV 306
               ||| ||  ::|            | :    |||   :: ||  |:|
Db      499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539


RESULT   6
P115_MYCHR
```

```
ID   P115_MYCHR     STANDARD;      PRT;   979 AA.
AC   P41508;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   P115 PROTEIN.
OS   Mycoplasma hyorhinis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2100;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91138990; PubMed=1825306;
RA   Notarnicola S.M., McIntosh M.A., Wise K.S.;
RT   "A Mycoplasma hyorhinis protein with sequence similarities to
RT   nucleotide-binding enzymes.";
RL   Gene 97:77-85(1991).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC       FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
CC       P115.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M34956; AAA25423.1; -.
DR   PIR; JQ0894; JQ0894.
KW   ATP-binding; Coiled coil.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   DOMAIN      169    224       COILED COIL (POTENTIAL).
FT   DOMAIN      231    400       COILED COIL (POTENTIAL).
FT   DOMAIN      569    821       COILED COIL (POTENTIAL).
FT   DOMAIN      884    912       ALA/ASP-RICH (DA-BOX).
SQ   SEQUENCE   979 AA;  110566 MW;  30D51C56B56280F4 CRC64;

  Query Match             5.8%;  Score 103;  DB 1;  Length 979;
  Best Local Similarity   19.6%;  Pred. No. 7;
  Matches   67;  Conservative   63;  Mismatches  126;  Indels   86;  Gaps   15;

Qy     11 VAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKE------------D 57
           :| :| ::::|:   |  :||  |      | ::    : |:     :||           |
Db    537 IAKILDHKYTIVSLEGD-LFRPGGTITGGSKLERTSILNYDIKIKEHTNTLKFAEDQIHD 595

Qy     58 LSWPGLAVGNLFHRPRATVM-------------VMVKGVNKLALPPGSVISYPLEN--A 101
          |      : |        :|:             :: : :| |  |     :    |:  :
Db    596 LKIKQQTIYNEIETVNSTIQQVKIEANSINSKLNILNEELNNLKLNASEIFKEQQEDQES 655

Qy    102 VPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMV----GKANSVFEDLSVTLRQLRNR 157
           :  | ||   :|    |   |    ::|    ::|:  :       ||  :  ::|   ||:| :
Db    656 LNLSFDSEKLNIEKQISTLT---IELNSKKDRLTNLISEQGKGETKKQELDAKLRKLNTQ 712

Qy    158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELA 217
                         :|::  |    |         :          || |  |  : : : |||:
Db    713 ----------HSDSITEQNRAKFL-------VEQNQKRLSEHYKLTLEAASEQYSLD-- 752

Qy    218 GLDEIGKRYGEDS--EQFRDASKILVDALQKFADDMYSLYGGNA--VVELVTVKS----- 268
          ||    |:  ||   :: ::   : ::|: :| :::    |          : || | ||
Db    753 -LDIEQARHFVDSLKKELKELGNVNLEAITEF-EEVNQRYQEKKQYIEELTTAKSKIEEA 810

Qy    269 ---FDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVF 307
               |  :| ||  |:                 ||     | |:::||
Db    811 ISDLDKIIINKTTEIV-----------NLV---NNEFNMVF 837

RESULT   7
MV10_MOUSE
```

```
ID   MV10_MOUSE     STANDARD;      PRT;  1004 AA.
AC   P23249;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROTEIN MOV-10.
GN   MOV10 OR GB110.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91117255; PubMed=1899287;
RA   Mooslehner K., Mueller U., Karls U., Hamann L., Harbers K.;
RT   "Structure and expression of a gene encoding a putative GTP-binding
RT   protein identified by provirus integration in a transgenic mouse
RT   strain.";
RL   Mol. Cell. Biol. 11:886-893(1991).
RN   [2]
RP   SEQUENCE OF 1-45 FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=94344088; PubMed=8065313;
RA   Hamann L., Bayer K.U., Jensen K., Harbers K.;
RT   "Interaction of several related GC-box- and GT-box-binding proteins
RT   with the intronic enhancer is required for differential expression of
RT   the gb110 gene in embryonal carcinoma cells.";
RL   Mol. Cell. Biol. 14:5786-5793(1994).
CC   -!- FUNCTION: IMPORTANT FUNCTION IN DEVELOPMENT AND/OR CONTROL
CC       OF CELL PROLIFERATION.
CC   -!- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X52574; CAA36803.1; -.
DR   EMBL; X75819; CAA53453.1; -.
DR   PIR; A39611; A39611.
DR   MGD; MGI:97054; Mov10.
DR   InterPro; IPR000886; -.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   ATP-binding; Helicase; Developmental protein.
FT   NP_BIND     525    532       ATP (BY SIMILARITY).
FT   SITE        646    649       DEAG BOX.
SQ   SEQUENCE   1004 AA;  113601 MW;  F0305D581846E758 CRC64;

  Query Match             5.8%;  Score 102.5;  DB 1;  Length 1004;
  Best Local Similarity   27.9%;  Pred. No. 7.9;
  Matches   56;  Conservative   19;  Mismatches   67;  Indels   59;  Gaps    9;

Qy    149 VTLRQLRNRLFQENSVLSS---------LPLNSLSRNNEVD---LLFLSELQVLHDISS 195
           : ||||   | |  |: ::            |        ||| ||     || | |||: |||
Db    299 ILLRQLLPTLLQGPSIFTAPKEVAEIKAQLETTLKSRNYEVKLRLLLHLEELQMEHDI-- 356

Qy    196 LLSRHKHLAK--------DHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKF 247
             ||  |                 | :| | :||: |: |           |     :|     |:
Db    357 ---RHYDLDSVPMTWDPVDQNPRLLTLEVPGVAESRPSVLRGDHLF----ALLSSETQQ- 408

Qy    248 ADDMYSLYGGNAVVELVTVK-SFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYS-- 304
           ||   :   |      |||  || || |||: :                      | :| ||  ::
Db    409 -DDPVTYKGFVHKVELDRVKLSFSTSLLSRFVD-------------GLTFKVNFTFNRQ 453

Qy    305 ----------VVFNMVLWIMI 315
                          :   ||| |:
Db    454 PLRVQHRALELTGRWVLWPML 474
```

```
RESULT   8
SIR4_YEAST
ID   SIR4_YEAST     STANDARD;      PRT;  1358 AA.
AC   P11978;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   REGULATORY PROTEIN SIR4 (SILENT INFORMATION REGULATOR 4).
GN   SIR4 OR STE9 OR ASD1 OR UTH2 OR YDR227W OR YD9934.12.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88142836; PubMed=3325825;
RA   Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT   "Functional domains of SIR4, a gene required for position effect
RT   regulation in Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:4441-4452(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95192063; PubMed=7885847;
RA   Davies C.J., Hutchison C.A. III;
RT   "Insertion site specificity of the transposon Tn3.";
RL   Nucleic Acids Res. 23:507-514(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE PROTEINS SIR1 THROUGH SIR4 CONTROL THE EXPRESSION OF
CC       MATING TYPE GENES RESIDENT AT LOCI OF EITHER END OF CHROMOSOME
CC       III. SIR3 AND SIR4 ASSOCIATE WITH THE C-TERMINUS OF RAP1 TO FORM A
CC       DNA-BINDING COMPLEX THAT INITIATES THE REPRESSION AT THE HM LOCI
CC       AND TELOMERES.
CC   -!- SUBUNIT: INTERACTS WITH RAP1 C-TERMINUS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M37249; AAA20881.1; -.
DR   EMBL; U13239; AAC33144.1; -.
DR   EMBL; Z48612; CAA88507.1; -.
DR   PIR; A29360; A29360.
DR   SGD; S0002635; SIR4.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW   Coiled coil.
FT   DOMAIN     1277   1347       COILED COIL (POTENTIAL).
FT   VARIANT     994    994       P -> L.
SQ   SEQUENCE   1358 AA;  152061 MW;  9C698765964F094E CRC64;

  Query Match             5.7%;  Score 100.5;  DB 1;  Length 1358;
  Best Local Similarity   24.1%;  Pred. No. 17;
  Matches   70;  Conservative   43;  Mismatches   92;  Indels   85;  Gaps   17;

Qy     38 PGERIPDVAALS----------MGFSVKEDLSWPGLAVGNLFHRPRATVMVMVKGVNKLA 87
          | ::||:    |            |:  :: | |  |:    |:|      |     :|
Db     23 PNDKIPEREEKSNEVKTPKIPLFTFAKSKNYSRPSTAIHTSPHQPSD---VKPTSHKQLQ 79

Qy     88 LPPGSVISYPLENAVPFS-LDSVANS-IHSLFSEETPVVLQLAPSEERVYMVGKANSVFE 145
           |  |  :     |: | | |: ::|| : ||    :|         |  |:    ::|:
Db     80 QPKSSPLKKNNYNSFPHSNLEKISNSKLLSLLRSKT--------SAGRI----ESNNPSH 127

Qy    146 DLSVTLRQLRNRLF-----QENSVLSS------LPLNSLSRNNEVDLLFLSELQVLHDIS 194
          | | :|      |      |: |  :|       :|:::   ||      |||        :
Db    128 DASRSLASFEQTAFSRHAQQQTSTFNSKPVRTIVPISTSQTNNS----FLS------GVK 177

Qy    195 SLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSL 254
          ||||  |     :|:| ::  : ||                :||      :|    |:| : |:
Db    178 SLLSEEK--IRDYSKEILGINLA------------NEQ-----PVLEKPLKKGSADI--- 215

Qy    255 YGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYS 304
            | :|: |    ||      |||  |: | |: |         |    || :|
Db    216 --GASVISLTKDKS-----IRKD-TVEEKKEEK-------LNIGKNFAHS 250

RESULT   9
NF1_HUMAN
ID   NF1_HUMAN      STANDARD;      PRT;  2839 AA.
AC   P21359;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).
GN   NF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.
RX   MEDLINE=92147138; PubMed=1783401;
RA   Marchuk D.A., Saulino A., Tavakkol R., Swaroop M., Wallace M.R.,
RA   Andersen L.B., Mitchell A.L., Gutmann D.H., Boguski M., Collins F.S.;
RT   "cDNA cloning of the type 1 neurofibromatosis gene: complete sequence
RT   of the NF1 gene product.";
RL   Genomics 11:931-940(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93090270; PubMed=1457041;
RA   Bernards A., Haase V.H., Murthy A.E., Menon A., Hannigan G.E.,
RA   Gusella J.F.;
RT   "Complete human NF1 cDNA sequence: two alternatively spliced mRNAs
RT   and absence of expression in a neuroblastoma line.";
RL   DNA Cell Biol. 11:727-734(1992).
RN   [3]
RP   SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.
RX   MEDLINE=90335969; PubMed=2116237;
RA   Xu G., O'Connell P., Viskochil D., Cawthon R., Robertson M.,
RA   Culver M., Dunn D., Stevens J., Gesteland R., White R., Weiss R.;
RT   "The neurofibromatosis type 1 gene encodes a protein related to GAP.";
RL   Cell 62:599-608(1990).
RN   [4]
RP   SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.
RX   MEDLINE=91029515; PubMed=2121370;
RA   Martin G.A., Viskochil D., Bollag G., McCabe P.C., Crosier W.J.,
RA   Haubruck H., Conroy L., Clark R., O'Connell P., Cawthon R.M.,
RA   Innis M., McCormick F.;
RT   "The GAP-related domain of the neurofibromatosis type 1 gene product
RT   interacts with ras p21.";
RL   Cell 63:843-849(1990).
RN   [5]
RP   SEQUENCE OF 1606-2709 FROM N.A., AND VARIANT PRO-1953.
RX   MEDLINE=90304909; PubMed=2114220;
RA   Cawthon R.M., Weiss R., Xu G., Viskochil D., Culver M., Stevens J.,
RA   Robertson M., Dunn D., Gesteland R., O'Connell P., White R.;
RT   "A major segment of the neurofibromatosis type 1 gene: cDNA sequence,
RT   genomic structure, and point mutations.";
RL   Cell 62:193-201(1990).
RN   [6]
RP   SEQUENCE OF 2230-2839 FROM N.A.
RX   MEDLINE=90319792; PubMed=2134734;
RA   Wallace M.R., Marchuk D.A., Andersen L.B., Letcher R., Odeh H.M.,
RA   Saulino A.M., Fountain J.W., Brereton A., Nicholson J., Mitchell A.L.,
RA   Brownstein B.H., Collins F.S.;
RT   "Type 1 neurofibromatosis gene: identification of a large transcript
RT   disrupted in three NF1 patients.";
RL   Science 249:181-186(1990).
RN   [7]
```

```
RP   ERRATUM.
RX   MEDLINE=91102559; PubMed=2125369;
RA   Wallace M.R., Marchuk D.A., Andersen L.B., Collins F.S.;
RL   Science 250:1749-1749(1990).
RN   [8]
RP   SEQUENCE OF 1168-1566 FROM N.A.
RX   MEDLINE=92019823; PubMed=1923522;
RA   Nishi T., Lee P.S., Oka K., Levin V.A., Tanase S., Morino Y.,
RA   Saya H.;
RT   "Differential expression of two types of the neurofibromatosis type 1
RT   (NF1) gene transcripts related to neuronal differentiation.";
RL   Oncogene 6:1555-1559(1991).
RN   [9]
RP   SEQUENCE OF 1371-1391 FROM N.A.
RX   MEDLINE=91109335; PubMed=8417346;
RA   Andersen L.B., Ballester R., Marchuk D.A., Chang E., Gutmann D.H.,
RA   Saulino A.M., Camonis J., Wigler M., Collins F.S.;
RT   "A conserved alternative splice in the von Recklinghausen
RT   neurofibromatosis (NF1) gene produces two neurofibromin isoforms,
RT   both of which have GTPase-activating protein activity.";
RL   Mol. Cell. Biol. 13:487-495(1993).
RN   [10]
RP   FUNCTION.
RX   MEDLINE=91029516; PubMed=2121371;
RA   Ballester R., Marchuk D., Boguski M.S., Saulino A., Letcher R.,
RA   Wigler M., Collins F.S.;
RT   "The NF1 locus encodes a protein functionally related to mammalian
RT   GAP and yeast IRA proteins.";
RL   Cell 63:851-859(1990).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=95072625; PubMed=7981724;
RA   Upadhyaya M., Shaw D.J., Harper P.S.;
RT   "Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis
RT   and polymorphisms in the NF1 gene.";
RL   Hum. Mutat. 4:83-101(1994).
RN   [12]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=96422425; PubMed=8825042;
RA   Hong Shen M., Harper P.S., Upadhyaya M.;
RT   "Molecular genetics of neurofibromatosis type 1 (NF1).";
RL   J. Med. Genet. 33:2-17(1996).
RN   [13]
RP   VARIANT GLU-1444.
RX   MEDLINE=92233464; PubMed=1568247;
RA   Li Y., Bollag G., Clark R., Stevens J., Conroy L., Fults D., Ward K.,
RA   Friedman E., Samowitz W., Robertson M., Bradley P., McCormick F.,
RA   White R., Cawthon R.;
RT   "Somatic mutations in the neurofibromatosis 1 gene in human tumors.";
RL   Cell 69:275-281(1992).
RN   [14]
RP   VARIANTS MET-2164 AND ASN-2192.
RX   MEDLINE=93258816; PubMed=1302608;
RA   Upadhyaya M., Shen M., Cherryson A., Farnham J., Maynard J.,
RA   Huson S.M., Harper P.S.;
RT   "Analysis of mutations at the neurofibromatosis 1 (NF1) locus.";
RL   Hum. Mol. Genet. 1:735-740(1992).
RN   [15]
RP   VARIANT HIS-1721--LEU-1733 DUPL.
RX   MEDLINE=93304433; PubMed=8317303;
RA   Tassabehji M., Strachan T., Sharland M., Colley A., Donnai D.,
RA   Harris R., Thakker N.;
RT   "Tandem duplication within a neurofibromatosis type 1 (NF1) gene exon
RT   in a family with features of Watson syndrome and Noonan syndrome.";
RL   Am. J. Hum. Genet. 53:90-95(1993).
RN   [16]
RP   VARIANT MET-991 DEL.
RX   MEDLINE=94108439; PubMed=7904209;
RA   Shen M.H., Harper P.S., Upadhyaya M.;
RT   "Neurofibromatosis type 1 (NF1): the search for mutations by PCR-
RT   heteroduplex analysis on Hydrolink gels.";
RL   Hum. Mol. Genet. 2:1861-1864(1993).
RN   [17]

RP   VARIANT NF1 ASN-2387--PHE-2388 DEL.
RX   MEDLINE=94362704; PubMed=8081387;
RA   Abernathy C.R., Colman S.D., Kousseff B.G., Wallace M.R.;
RT   "Two NF1 mutations: frameshift in the GAP-related domain, and loss of
RT   two codons toward the 3' end of the gene.";
RL   Hum. Mutat. 3:347-352(1994).
RN   [18]
RP   VARIANT NF1 ALA-2631.
RX   MEDLINE=96091873; PubMed=8544190;
RA   Upadhyaya M., Maynard J., Osborn M., Huson S.M., Ponder M.,
RA   Ponder B.A.J., Harper P.S.;
RT   "Characterisation of germline mutations in the neurofibromatosis type
RT   1 (NF1) gene.";
RL   J. Med. Genet. 32:706-710(1995).
RN   [19]
RP   VARIANT NF1 ARG-629.
RX   MEDLINE=96431167; PubMed=8834249;
RA   Gasparini P., D'Agruma L., de Cillis G.P., Balestrazzi P.,
RA   Mingarelli R., Zelante L.;
RT   "Scanning the first part of the neurofibromatosis type 1 gene by RNA-
RT   SSCP: identification of three novel mutations and of two new
RT   polymorphisms.";
RL   Hum. Genet. 97:492-495(1996).
RN   [20]
RP   VARIANT LS ARG-1035.
RX   MEDLINE=96400960; PubMed=8807336;
RA   Wu R., Legius E., Robberecht W., Dumoulin M., Cassiman J.-J.,
RA   Fryns J.-P.;
RT   "Neurofibromatosis type 1 gene mutation in a patient with features of
RT   LEOPARD syndrome.";
RL   Hum. Mutat. 8:51-56(1996).
RN   [21]
RP   VARIANTS NF1 ARG-844 AND PRO-898.
RX   MEDLINE=97295087; PubMed=9150739;
RA   Maynard J., Krawczak M., Upadhyaya M.;
RT   "Characterization and significance of nine novel mutations in exon 16
RT   of the neurofibromatosis type 1 (NF1) gene.";
RL   Hum. Genet. 99:674-676(1997).
RN   [22]
RP   VARIANT NF1 ARG-1952.
RX   MEDLINE=97255969; PubMed=9101300;
RA   Hudson J., Wu C.L., Tassabehji M., Summers E.M., Simon S., Super M.,
RA   Donnai D., Thakker N.;
RT   "Novel and recurrent mutations in the neurofibromatosis type 1 (NF1)
RT   gene.";
RL   Hum. Mutat. 9:366-367(1997).
RN   [23]
RP   VARIANT NF1 TRP-1611.
RX   MEDLINE=97442280; PubMed=9298829;
RA   Upadhyaya M., Maynard J., Osborn M., Harper P.S.;
RT   "Six novel mutations in the neurofibromatosis type 1 (NF1) gene.";
RL   Hum. Mutat. 10:248-250(1997).
RN   [24]
RP   VARIANT NF1 PRO-1446.
RX   MEDLINE=99235555; PubMed=10220145;
RA   Peters H., Hess D., Fahsold R., Schuelke M.;
RT   "A novel mutation (1442SP in the CAP-region of the NF1 gene detected by
RT   temperature gradient gel electrophoresis (TGGE).";
RL   Hum. Mutat. 13:337-337(1999).
CC   -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER
CC       AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE
CC       A REGULATOR OF RAS ACTIVITY.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I AND II (SHOWN HERE); ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DISEASE: THIS PROTEIN IS ASSCCIATED WITH TYPE 1 NEUROFIBROMATOSIS
CC       (NF1) (ALSO CALLED VON RECKLINGHAUSEN SYNDROME), THE MOST FREQUENT

 Query Match              5.7%;  Score 100.5;  DB 1;  Length 2839;
 Best Local Similarity   22.2%;  Pred. No. 50;
 Matches   86;  Conservative   60;  Mismatches  140;  Indels  101;  Gaps   18;

Qy      13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
           | ||:  : : :  :|   :|| :          :| |   ||   :::: | |  :|:
```

```
Db   2021 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 2080

Qy     66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                  | |:|:   | | :           |::: |    | || :     | | |||
Db   2081 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLVINIIHS 2132

Qy    116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
          |       |||||  ||:|: :|     : |::    : |      :   |   |:| |    |
Db   2133 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 2192

Qy    165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
             :  | ||     |  || :  |     : || :    |  :   ||:       ::| |    |
Db   2193 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 2250

Qy    218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS--------------------- 253
               | ||  :|:  :  |  || |    |:     | |:
Db   2251 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIEATVIALTKLQPLLNKD 2308

Qy    254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP----------- 292
                 :    ||::|  | : |  |:|: :    |:: :   |  ||
Db   2309 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2368

Qy    293 ---------YNLAYKYNFEYSVVFNMV 310
                   | :  || :::| :::
Db   2369 WHCKQMDHFVGLNFNSNFNFALVGHLL 2395
```

```
RESULT  10
SYD_MYCCA
ID   SYD_MYCCA      STANDARD;      PRT;   483 AA.
AC   Q48979;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12) (ASPARTATE--TRNA LIGASE)
DE   (ASPRS) (FRAGMENT).
GN   ASPS.
OS   Mycoplasma capricolum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Entomoplasmataceae.
OX   NCBI_TaxID=2095;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27343 / KID;
RX   MEDLINE=96059641; PubMed=7476192;
RA   Bork P., Ouzounis C., Casari G., Schneider R., Sander C.,
RA   Dolan M., Gilbert W., Gillevet P.M.;
RT   "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
RT   its physiology.";
RL   Mol. Microbiol. 16:955-967(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +
CC       PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z33048; CAA83720.1; -.
DR   InterPro; IPR002106; -.
DR   InterPro; IPR002309; -.
DR   Pfam; PF00152; tRNA-synt_2; 1.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; PARTIAL.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   NON_TER       1      1
SQ   SEQUENCE   483 AA;  56298 MW;  A03F5B0CEA31087E CRC64;
```

```
  Query Match             5.6%;  Score 99;  DB 1;  Length 483;
  Best Local Similarity   25.8%;  Pred. No. 5;
  Matches   62;  Conservative   38;  Mismatches   84;  Indels   56;  Gaps   14;

Qy     76 VMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVY 135
            ::|  | |||          :|:  :|  |    |   | |          || || ||:
Db     31 IVVKRKSVNK-----ELITGEIEVIVKDLL--VINK-----SELTPFVLE--------- 68

Qy    136 MVGKANSVFEDLSVTLR--QLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHD- 192
              :| ||  :| |     || :: | | :: :  :| : ||    || || |:: :
Db     69 ---NDVNVNEDTRLTYRYLDLRRQVMQNNLIIRA-KINHIIRNYLTDLNFL-EVETPYFA 123

Qy    193 -----------ISSLLSRHKHLAKDHSPDLYS--LELAGLDEIGKRYGEDSEQFRDASKI 239
                     : | |:::|  |    || |:   | ::|:|     || :     |||   :
Db    124 KSTPEGARHFLVPSRLNKNKFYALPQSPQLFKQLLMISGID----RYYQIVRCFRD-EDL 178

Qy    240 LVDALQKFAD-DMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYK 298
             :|   :|   |:   :   :          |      |||:|       :: :  |     |:||
Db    179 RIDRQPEFTQLDLEMSFATSE-----DVMQISESLIKKILKEVKNFEIKEPL--LRLSYK 231
```

```
RESULT  11
MYSB_HUMAN
ID   MYSB_HUMAN     STANDARD;      PRT;  1935 AA.
AC   P12883; Q14904; Q16579;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN   MYH7 OR MYHCB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91065634; PubMed=2249844;
RA   Jaenicke T., Diederich K.W., Haas W., Schleich J., Lichter P.,
RA   Pfordt M., Bach A., Vosberg H.P.;
RT   "The complete sequence of the human beta-myosin heavy chain gene and
RT   a comparative analysis of its product.";
RL   Genomics 8:194-206(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90301496; PubMed=2362820;
RA   Liew C.-C., Sole M.J., Yamauchi-Takihara K., Kellam B.,
RA   Anderson D.H., Lin L., Liew J.;
RT   "Complete sequence and organization of the human cardiac beta-myosin
RT   heavy chain gene.";
RL   Nucleic Acids Res. 18:3647-3651(1990).
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=89264452; PubMed=2726733;
RA   Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT   "Characterization of human cardiac myosin heavy chain genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN   [4]
RP   ERRATUM.
RA   Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL   Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN   [5]
RP   SEQUENCE OF 1310-1935 FROM N.A.
RX   MEDLINE=86176778; PubMed=2421254;
RA   Saez L., Leinwand L.A.;
RT   "Characterization of diverse forms of myosin heavy chain expressed in
RT   adult human skeletal muscle.";
RL   Nucleic Acids Res. 14:2951-2969(1986).
RN   [6]
RP   REVISIONS.
RA   Leinwand L.A.;
RL   Submitted (MAR-1988) to the EMBL/GenBank/DDBJ databases.
```

RN [7]
RP SEQUENCE OF 1410-1935 FROM N.A.
RX MEDLINE=88299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
RT beta-form myosin heavy chain complementary DNA clones. Regulation of
RT expression during development and pressure overload in human
RT atrium.";
RL J. Clin. Invest. 82:524-531(1988).
RN [8]
RP SEQUENCE OF 785-1935 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
RT myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [9]
RP SEQUENCE OF 1393-1935 FROM N.A.
RX MEDLINE=87192738; PubMed=3032766;
RA Jandreski M.A., Liew C.-C.;
RT "Construction of a human ventricular cDNA library and
RT characterization of a beta myosin heavy chain cDNA clone.";
RL Hum. Genet. 76:47-53(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=96039076; PubMed=8533830;
RA Arai S., Matsuoka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
RA Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
RA Hosoda S., Momma K.;
RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
RT hypertrophic cardiomyopathy.";
RL Am. J. Med. Genet. 58:267-276(1995).
RN [11]
RP VARIANTS CMH1 GLU-256 AND ARG-741.
RX MEDLINE=93248216; PubMed=8483915;
RA Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;
RT "Missense mutations in the beta-myosin heavy-chain gene cause central
RT core disease in hypertrophic cardiomyopathy.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
RN [12]
RP VARIANT CMH1 GLN-403.
RX MEDLINE=90367131; PubMed=1975517;
RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
RA McKenna W., Seidman C.E., Seidman J.G.;
RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
RT cardiac myosin heavy chain gene missense mutation.";
RL Cell 62:999-1006(1990).
RN [13]
RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
RX MEDLINE=92204193; PubMed=1552912;
RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,
RA Seidmann C.E., Seidmann J.G.;
RT "Characteristics and prognostic implications of myosin missense
RT mutations in familial hypertrophic cardiomyopathy.";
RL New Engl. J. Med. 326:1108-1114(1992).
RN [14]
RP VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
RX MEDLINE=94070863; PubMed=8250038;
RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,
RA McKenna W., Seidman J.G., Seidman C.E.;
RT "Independent origin of identical beta cardiac myosin heavy-chain
RT mutations in hypertrophic cardiomyopathy.";
RL Am. J. Hum. Genet. 53:1180-1185(1993).
RN [15]
RP VARIANTS CMH1 GLN-403 AND VAL-908.
RX MEDLINE=92346810; PubMed=1638703;
RA Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;
RT "Differences in clinical expression of hypertrophic cardiomyopathy
RT associated with two distinct mutations in the beta-myosin heavy chain
RT gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
RL Circulation 86:345-352(1992).

RN [16]
RP VARIANTS CMH1 LEU-403 AND TRP-403.
RX MEDLINE=94075629; PubMed=8254035;
RA Dausse E., Komajda M., Fetler L., Dubourg O., Dufour C., Carrier L.,
RA Wisnewsky C., Bercovici J., Hengstenberg C., Al-Mahdawi S.;
RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
RT identification of a hot spot for mutations in the beta-myosin heavy
RT chain gene.";
RL J. Clin. Invest. 92:2807-2813(1993).
RN [17]
RP VARIANT CMH1 TRP-403.
RX MEDLINE=94093568; PubMed=8268932;
RA Moolman J.C., Brink P.A., Corfield V.A.;
RT "Identification of a new missense mutation at Arg403, a CpG mutation
RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
RT hypertrophic cardiomyopathy.";
RL Hum. Mol. Genet. 2:1731-1732(1993).
RN [18]
RP VARIANT CMH1 ASN-615.
RX MEDLINE=93038688; PubMed=1417858;
RA Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
RT "Novel missense mutation in cardiac beta myosin heavy chain gene
RT found in a Japanese patient with hypertrophic cardiomyopathy.";
RL Biochem. Biophys. Res. Commun. 188:379-387(1992).
RN [19]
RP VARIANT CMH1 GLY-778.
RX MEDLINE=93343938; PubMed=8343162;
RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
RT "A missense mutation of cardiac beta-myosin heavy chain gene linked
RT to familial hypertrophic cardiomyopathy in affected Japanese
RT families.";
RL Biochem. Biophys. Res. Commun. 194:791-798(1993).
RN [20]
RP VARIANT CMH1 VAL-908.
RX MEDLINE=93168485; PubMed=8435239;
RA Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P.,
RA Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
RT "Identification of a mutation in the beta cardiac myosin heavy chain
RT gene in a family with hypertrophic cardiomyopathy.";
RL Br. Heart J. 69:136-141(1993).
RN [21]
RP VARIANT CMH1 TRP-719.
RX MEDLINE=95179132; PubMed=7874131;
RA Greve G., Bachinski L., Friedman D.L., Czernuszewicz G., Anan R.,
RA Towbin J., Seidman C.E., Roberts R.;
RT "Isolation of a de novo mutant myocardial beta MHC protein in a
RT pedigree with hypertrophic cardiomyopathy.";
RL Hum. Mol. Genet. 3:2073-2075(1994).
RN [22]
RP VARIANTS CMH1 CYS-513; ARG-716, AND TRP-719.
RX MEDLINE=94110336; PubMed=8282798;
RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
RA Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,
RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
RT mutations that cause familial hypertrophic cardiomyopathy.";
RL J. Clin. Invest. 93:280-285(1994).
RN [23]
RP VARIANT CMH1 THR-797.
RX MEDLINE=96047159; PubMed=7581410;
RA Moolman J.C., Brink P.A., Corfield V.A.;
RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
RL Hum. Mutat. 6:197-198(1995).
RN [24]
RP VARIANT CMH1 CYS-453.
RX MEDLINE=96205901; PubMed=8655135;
RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
RA Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
RT "Malignant familial hypertrophic cardiomyopathy in a family with a
RT 453Arg-->Cys mutation in the beta-myosin heavy chain gene:
RT coexistence of sudden death and end-stage heart failure.";
RL Hum. Genet. 97:585-590(1996).

```
RN   [25]
RP   VARIANTS CMH1 THR-349 AND TRP-719.
RX   MEDLINE=98204402; PubMed=9544842;
RA   Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T.,
RA   Dohlemann C., Vosberg H.-P.;
RT   "A high risk phenotype of hypertrophic cardiomyopathy associated with

  Query Match             5.6%;  Score 98.5;  DB 1;  Length 1935;
  Best Local Similarity   24.7%;  Pred. No. 40;
  Matches   46;  Conservative   39;  Mismatches   76;  Indels   25;  Gaps    8;

Qy     115 SLFSEETPVVLQ-------LAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSS 167
           || |: : ||       || :|||  : | | :  |  :::: || | : :
Db     879 SLLQEKNDLQLQVQAEQDNLADAEERCDQLIK-NKI--QLEAKVKEMNERLEDEEEMNAE 935

Qy     168 LPLNSLSRNNEVDLLFLSELQVLHDISSL------LSRHKHLAKDHSPDLYSLELAGLDE 221
           |     :|     |||:   ||  |       : : ||  ::   :| : |:||||
Db     936 LTAKKRKLEDE-----CSELK--RDIDDLELTLAKVEKEKHATENKVKNL-TEEMAGLDE 987

Qy     222 IGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTIL 281
           |  :  :: :  ::| :  :| ||  | : :|    :|   |    : || :: :  :
Db     988 IIAKLTKEKKALQEAHQQALDDLQAEEDKVNTLTKAKVKLE-QQVDDLEGSLEQEKKVRM 1046

Qy     282 EAKQAK 287
           : ::||
Db    1047 DLERAK 1052

RESULT 12
SECA_PORPU
ID   SECA_PORPU     STANDARD;      PRT;   884 AA.
AC   P51381;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PREPROTEIN TRANSLOCASE SECA SUBUNIT.
GN   SECA.
OS   Porphyra purpurea.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX   NCBI_TaxID=2787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AVONPORT;
RA   Reith M.E., Munholland J.;
RT   "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT   genome.";
RL   Plant Mol. Biol. Rep. 13:333-335(1995).
CC   -!- FUNCTION: INVOLVED IN PROTEIN EXPORT AND IMPORT. SECA HAS A
CC       CENTRAL ROLE IN COUPLING THE HYDROLYSIS OF ATP TO THE TRANSFER OF
CC       PRE-SECRETORY PROTEINS ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC   -!- SIMILARITY: BELONGS TO THE SECA FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U38804; AAC08267.1; -.
DR   Mendel; 10360; PORpu;secA;1.
DR   InterPro; IPR000185; -.
DR   Pfam; PF01043; SecA_protein; 1.
DR   PRINTS; PR00906; SECA.
DR   PROSITE; PS01312; SECA; 1.
KW   Protein transport; ATP-binding; Chloroplast; Translocation; Transport.
FT   NP_BIND      98    105       ATP (BY SIMILARITY).
SQ   SEQUENCE   884 AA;  101325 MW;  A415846D12B90B2B CRC64;

  Query Match             5.5%;  Score 97.5;  DB 1;  Length 884;
  Best Local Similarity   22.8%;  Pred. No. 15;
  Matches   61;  Conservative   47;  Mismatches   92;  Indels   67;  Gaps   13;

Qy      94 ISYPLENAVPFSLDSVANSI-----------HSLFSEETPVVLQLAPSEERVYMVGKAN 141
           | : | || | ::: :: |              ::    | ::|  ||    |:  |
Db     451 IPHSLLNAKPENVEKESDIIAQAGRQSSVTIATNMAGRGTDIILGGNPS----YI---AK 503

Qy     142 SVFEDLSVTLRQLRNRL-FQENSVLSSLPLNSLSRNNEVDLLF--LSELQVLHDIS---- 194
           |:  || :    ::| |  |: |: :   :| : :: | ||::  || |::    ||
Db     504 SILIDLLIKKISVQNNLKLQQLSLKTQYCINQILKSLEDDLIYANLSVLELEKKISIACE 563

Qy     195 -SLLSRHKHLAKDHSPDL-------------YSLELAGLDEIGKRYGED---SEQFR-- 234
              :||:  :    :  |                |   :  ||  ||     |   | |
Db     564 QVAISRNLEIQLRKAYQLIFQEYENIFSQEKKYVAQAGGLHVIGTERHESRRIDNQLRGR 623

Qy     235 -------DASKILVDALQKFADDMYSLYGGNAVVELVTVKSFD------TSLIRKTRTIL 281
                  :|:  :        |::  ::|||  : :|:    : |       ::|: |:   |
Db     624 AGRQGDPGSSRFFL----SIEDNLLRIFGGNKIADLMQALNVDDDTPMESTLLSKS---L 676

Qy     282 EAKQAKNPASPYNLAYKYNFEYSVVFN 308
           || | |  |  |:   |  |||  ||
Db     677 EAAQKKVEAYFYD-TRKQVFEYDQVLN 702

RESULT 13
NF1_MOUSE
ID   NF1_MOUSE      STANDARD;      PRT;  2841 AA.
AC   Q04690; Q61956; Q61957;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).
GN   NF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Brain;
RX   MEDLINE=93357730; PubMed=8353485;
RA   Bernards A., Snijders A.J., Hannigan G.E., Murthy A.E., Gusella J.F.;
RT   "Mouse neurofibromatosis type 1 cDNA sequence reveals high degree of
RT   conservation of both coding and non-coding mRNA segments.";
RL   Hum. Mol. Genet. 2:645-650(1993).
RN   [2]
RP   SEQUENCE OF 1178-1555 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=95047432; PubMed=7958951;
RA   Mantani A., Makasugi S., Yokota Y., Abe K., Ushio Y., Yamamura K.;
RT   "A novel isoform of the neurofibromatosis type-1 mRNA and a switch of
RT   isoforms during murine cell differentiation and proliferation.";
RL   Gene 148:245-251(1994).
RN   [3]
RP   SEQUENCE OF 1950-2568 FROM N.A.
RX   MEDLINE=90384569; PubMed=2169593;
RA   Buchberg A.M., Cleveland L.S., Jenkins N.A., Copeland N.G.;
RT   "Sequence homology shared by neurofibromatosis type-1 gene and IRA-1
RT   and IRA-2 negative regulators of the RAS cyclic AMP pathway.";
RL   Nature 347:291-294(1990).
CC   -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER
CC       AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE
CC       A REGULATOR OF RAS ACTIVITY.
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; I, II (SHOWN HERE), III AND
CC       IV; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: TYPE I IS EXPRESSED PREDOMINANTLY IN BRAIN,
CC       SPINAL CORD AND TESTIS. TYPE II IS EXPRESSED PREDOMINANTLY IN
CC       ADRENAL GLAND, KIDNEY, OVARY AND LUNG. TYPE III IS EXPRESSED
CC       PREDOMINANTLY IN ADRENAL GLAND AND TYPE IV IS EXPRESSED
CC       MAINLY IN THE TESTIS.
CC   -!- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L10369; AAA39806.1; -.
DR   EMBL; L10367; AAA39806.1; JOINED.
DR   EMBL; L10368; AAA39806.1; JOINED.
DR   EMBL; L10370; AAA68132.1; -.
DR   EMBL; X54924; CAA38690.1; -.
DR   EMBL; D30730; BAA06395.1; -.
DR   EMBL; D30731; BAA06396.1; -.
DR   MGD; MGI:97306; Nf1.
DR   InterPro; IPR001936; -.
DR   Pfam; PF00616; RasGAP; 1.
DR   PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR   PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW   GTPase activation; Alternative splicing.
FT   DOMAIN     1237   1453       RAS-GAP.
FT   VARSPLIC   1373   1393       MISSING (IN ISOFORM I AND ISOFORM IV).
FT   VARSPLIC   1394   1406       VVSQRFPQNSIGA -> VPKSSCFSCLNNRWLASASLRT
FT                                ASVP (IN ISOFORM III AND ISOFORM IV).
FT   VARSPLIC   1407   2841       MISSING (IN ISOFORM III AND ISOFORM IV).
SQ   SEQUENCE   2841 AA;  319591 MW;  06650BDCDC531002 CRC64;

  Query Match              5.5%;  Score 97.5;  DB 1;  Length 2841;
  Best Local Similarity   22.0%;  Pred. No. 83;
  Matches   85;  Conservative   61;  Mismatches  140;  Indels  101;  Gaps   18;

Qy       13 GVLGNEFSILKSPGSVVFRNGNWPIPGE-------RIPDVAALSMGFSVKEDLSWPGLAV 65
             | ||:  : :  :  :|    :|| :              :|  |    ||     :::: |  |   :|:
Db     2023 GGLGSIKAEVMADTAVALASGNVKLVSSKVIGRMCKIIDKTCLSPTPTLEQHLMWDDIAI 2082

Qy       66 GNLFHRPRATVMVMVKGVNKLALPPG-----SVISYPLENAVPFSLDS-----VANSIHS 115
                    |  |:|:   |  | :           |::: |     | || :       : | |||
Db     2083 -------LARYMLMLSFNNSLDVAAHLPYLFHVVTF-LVATGPLSLRASTHGLLINIIHS 2134

Qy      116 L-------FSEETPVVLQLAPSE---ERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSV 164
            |        |||||  ||:|: :|     : |::      : |         :     |   |:| |      |
Db     2135 LCTCSQLHFSEETKQVLRLSLTEFSLPKFYLLFGISKVKSAAVIAFRSSYRDRSFSPGSY 2194

Qy      165 -LSSLPLNSLSRNNEVDLLFLSELQVLHDISSL--LSRHKHLAK----DHSPDLYSLELA 217
              :  | ||         |   || : |     :    : ||  :    |  :    ||:      ::| |      |
Db     2195 ERETFALTSLETVTEA-LLEIME-ACMRDIPTCKWLDQWTELAQRFAFQYNPSLQPRALV 2252

Qy      218 GLDEIGKR--YGEDSEQFRDASKILVDALQKFADDMYS---------------------- 253
                |  ||   :|:   :    |    ||  |     |:        |  |:
Db     2253 VFGCISKRVSHGQIKQIIRILSKALESCLK--GPDTYNSQVLIESTVIALTKLQPLLNKD 2310

Qy      254 -------LYGGNAVVEL--VTVKSFDTSLIRKTRTILEAKQAKNPASP------------ 292
                    :     ||::|   |  :  |     |:|:  :        |::  :     |     ||
Db     2311 SPLHKALFWVAVAVLQLDEVNLYSAGTALLEQNLHTLDSLRIFNDKSPEEVFMAIRNPLE 2370

Qy      293 ---------YNLAYKYNFEYSVVFNMV 310
                     |  :   ||  :::|  :::
Db     2371 WHCKQMDHFVGLNFNSNFNFALVGHLL 2397

RESULT  14
ORC3_MOUSE
ID   ORC3_MOUSE     STANDARD;      PRT;   715 AA.
AC   Q9JK30;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ORIGIN RECOGNITION COMPLEX SUBUNIT 3 (ORIGIN RECOGNITION COMPLEX
DE   SUBUNIT LATHEO).
GN   ORC3L OR ORC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=10702681;
RA   Springer J., Nanda I., Hoehn K., Schmid M., Grummt F.;
RT   "Identification and chromosomal localization of murine ORC3, a new
RT   member of the mouse origin recognition complex.";
RL   Cytogenet. Cell Genet. 87:245-251(1999).
CC   -!- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
CC       BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
CC       REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
CC       THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN
CC       ATP-DEPENDENT MANNER.
CC   -!- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE ORC3 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ132830; CAB76399.1; -.
DR   MGD; MGI:1354944; Orc3.
KW   DNA replication; Nuclear protein; DNA-binding.
SQ   SEQUENCE   715 AA;  82341 MW;  1E915A719A2E33EB CRC64;

  Query Match              5.5%;  Score 97;  DB 1;  Length 715;
  Best Local Similarity   22.0%;  Pred. No. 12;
  Matches   80;  Conservative   45;  Mismatches  116;  Indels  122;  Gaps   17;

Qy       67 NLFHRPRATVMVMVKGVNKLAL------PPGSVISYPLENAVP----FSLDSVANSIHSL 116
             |:|      ::    :||:  ||:|           | ||:    |    |        ||:      |  |   |
Db      318 NIFLYHDFSIQSFIKGI-KLSLLEHFYSQPLSVLCCDLSEAKKRVNVFSVSQCEN-IRRL 375

Qy      117 -----FSEETPVVLQLAPSEERVYMVGKANSVFEDLSV-------TLRQLRNRLFQENSV 164
                  : |   |:   |:|          ::  |    |: ||| |              || |  |
Db      376 PSFRRYVENQPLGKQVALLTNETFLKEKTQSLLEDLHVYHINYFLVLRCLHN-------F 428

Qy      165 LSSLPLNSLSRNNEVDLLFLSELQV----LHDISSLLSRHKHLAKDHSPDLYSLELAGLD 220
            ||||      |  |   ::   |:  :  |:     |  |      :  ||||        :    ||
Db      429 TSSLPKYPLGR--QIRELYCTCLEKKIWDSEEYKSALQLLRMLAKDELVSILQRCIEVLD 486

Qy      221 -----EIG------------KRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVE 262
                  ::|               :       ||::   ||          |||     |:|  |          :::|
Db      487 SSTEKQLGNTTQKIKDFLTQFQNLDADSKEEEDACGSQPKGLQK--TDLYHLQ--KSLLE 542

Qy      263 L----------------VTVKSFDTSLIR-----------------KTRTILEAKQA-- 286
            :                  | :|    :|:|                             |:  |     |
Db      543 MKELRRTKKPTKFEMLRENVMNFIDNLVRDYLLPPESQPLHEVVYFSAANTLREHLNAAP 602

Qy      287 ----------------------KNPASPYNLAYKYNFEYSVVFNMVLWIMIALALAVI 322
                                  | |     :|||  :  | ||  |::  |:|  |      | |  |   :
Db      603 RIALHTALNNPYYYLKNEELEGCIPNTAPDICIAYKLHLECSLLINLVDW---AEAFATV 659

Qy      323 ITS 325
             :|:
Db      660 VTA 662

RESULT  15
SUS1_TULGE
ID   SUS1_TULGE     STANDARD;      PRT;   805 AA.
AC   Q41608;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   SUCROSE SYNTHASE 1 (EC 2.4.1.13) (SUCROSE-UDP GLUCOSYLTRANSFERASE 1).
```

```
OS   Tulipa gesneriana.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Liliales; Liliaceae; Tulipa.
OX   NCBI_TaxID=13306;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. APELDOORN;
RA   Balk P.A., de Boer A.D.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND
CC       FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE = UDP + SUCROSE.
CC   -!- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X96938; CAA65639.1; -.
DR   InterPro; IPR000368; -.
DR   InterPro; IPR001296; -.
DR   Pfam; PF00534; Glycos_transf_1; 1.
DR   Pfam; PF00862; Sucrose_synth; 1.
KW   Transferase; Glycosyltransferase; Multigene family.
SQ   SEQUENCE   805 AA;  92254 MW;  B749D8953C9338F0 CRC64;

  Query Match             5.5%;  Score 96.5;  DB 1;  Length 805;
  Best Local Similarity   22.9%;  Pred. No. 16;
  Matches   59;  Conservative   39;  Mismatches  101;  Indels   59;  Gaps   11;

Qy     60 WPGLAVGNLFHRPRATVMVMVK-GVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFS 118
          |  ||:     |||  |   |:  ||:||:   |         : |  | |  |  |  |
Db     84 WVALAI-----RPRPGVWEYVRVNVNELAVEECS-------EYLKFKEDLVDRSSQSNFV 131

Qy    119 EETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNN- 177
           |    :   |    |     : |:   :    | | ::|| :   |  | || |  :|
Db    132 LE----MDFEPFNANVPRPSLSKSIGNGVQFLNRHLSSKLFHDKESLYPL-LNFLREHNY 186

Qy    178 -EVDLLFLSELQVLHDISSLLSRHKH----LAKD--HSPDLYSLELAGLDEIGKRYGEDS 230
              |:    || |  : : | :       ::||  :|   :| :: ||:   | :|: :
Db    187 KGTTLMLNDRLQSLSALQTALRKADRYLLSISKDTPYSEFNHSFQVLGLE---KGWGDTA 243

Qy    231 EQFRDASKILVD--------ALQKFADDM----------------------YSLYGGNAV 260
           :  :   :|:|         |:||   :                      |   ||  |
Db    244 SRVSENIHLLLDLLEAPDPSTLEKFLGTIPMVFNVVILSPHGYFAQANVLGYPDTGGQVV 303

Qy    261 VELVTVKSFDTSLIRKTR 278
            |  |:: :| :: | :
Db    304 YILDQVRALETEMLLKIK 321


Search completed: July  3, 2001, 16:19:44
Job time: 141 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:17:48 ; Search time 28.21 Seconds
(without alignments)
1641.502 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVFVVLLALVAGVLGNEFS.....MDPGYDSIIYRMTNQKIRMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1759	99.8	350	4	Q9H2P8	Q9h2p8 homo sapien
2	1667	94.6	330	4	Q9UG89	Q9ug89 homo sapien
3	327	18.5	320	5	Q9VHG4	Q9vhg4 drosophila
4	192.5	10.9	324	5	Q9XTY9	Q9xty9 caenorhabdi
5	121.5	6.9	364	10	Q9ZSJ7	Q9zsj7 arabidopsis
6	114	6.5	663	3	O74359	O74359 schizosacch
7	109.5	6.2	408	5	Q22736	Q22736 caenorhabdi
8	107.5	6.1	251	5	Q9N598	Q9n598 caenorhabdi
9	107	6.1	1112	10	Q9FH47	Q9fh47 arabidopsis
10	105	6.0	458	5	Q93377	Q93377 caenorhabdi
11	104	5.9	412	2	O25885	O25885 helicobacte
12	103.5	5.9	732	2	O68717	O68717 yersinia pe
13	103.5	5.9	732	2	Q9RI12	Q9ri12 yersinia pe
14	103.5	5.9	925	10	Q9SIX4	Q9six4 arabidopsis
15	103	5.8	470	2	Q9ZDN3	Q9zdn3 rickettsia
16	102.5	5.8	1380	5	Q9VIU8	Q9viu8 drosophila
17	102	5.8	403	4	Q9NX71	Q9nx71 homo sapien
18	101	5.7	412	2	Q9RJG7	Q9rjg7 helicobacte
19	101	5.7	689	3	P87169	P87169 schizosacch
20	100.5	5.7	2820	11	P97526	P97526 rattus norv
21	99.5	5.6	861	5	Q9N519	Q9n519 caenorhabdi
22	98.5	5.6	471	3	Q9P723	Q9p723 neurospora -
23	98.5	5.6	728	2	Q56921	Q56921 yersinia en
24	98.5	5.6	1038	6	Q28699	Q28699 oryctolagus
25	98.5	5.6	1935	4	Q9H1D5	Q9h1d5 homo sapien
26	98.5	5.6	2763	13	Q9YGV2	Q9ygv2 fugu rubrip
27	98	5.6	561	1	Q9V2L7	Q9v2l7 pyrococcus
28	98	5.6	864	10	Q9T033	Q9t033 arabidopsis
29	98	5.6	980	14	P90478	P90478 feline sync
30	97.5	5.5	758	1	Q9UY62	Q9uy62 pyrococcus
31	97.5	5.5	982	14	O56861	O56861 feline foam
32	97.5	5.5	1348	5	Q9VAD1	Q9vad1 drosophila
33	96.5	5.5	533	1	Q9YCP2	Q9ycp2 aeropyrum p
34	96.5	5.5	574	10	Q9FY55	Q9fy55 arabidopsis
35	96.5	5.5	729	2	O85239	O85239 yersinia en
36	96.5	5.5	1062	2	P95422	P95422 pseudomonas
37	96.5	5.5	1062	2	Q9IUY8	Q9iuy8 pseudomonas
38	96.5	5.5	1935	6	Q9GKR1	Q9gkr1 sus scrofa
39	96	5.4	390	2	Q9JZF1	Q9jzf1 neisseria m
40	95.5	5.4	1020	2	Q9KLN4	Q9kln4 vibrio chol
41	95.5	5.4	1209	5	Q93620	Q93620 caenorhabdi
42	95.5	5.4	2471	14	Q9YTK3	Q9ytk3 ateline her
43	95	5.4	2253	13	P70012	P70012 xenopus lae
44	94.5	5.4	769	4	Q9H4E9	Q9h4e9 homo sapien
45	94	5.3	1778	3	Q9UTK4	Q9utk4 schizosacch

ALIGNMENTS

RESULT 1
Q9H2P8
ID Q9H2P8 PRELIMINARY; PRT; 350 AA.
AC Q9H2P8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HT028.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HYPOTHALAMUS;
RA Yang Y., Xu X., Gao G., Xiao H., Chen Z., Han Z.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF248966; AAG44564.1; -.
SQ SEQUENCE 350 AA; 39036 MW; 840A989ACC4E8DE8 CRC64;

Query Match 99.8%; Score 1759; DB 4; Length 350;
Best Local Similarity 99.7%; Pred. No. 5.8e-126;
Matches 349; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAVFVVLLALVAGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSW 60

Qy      61 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEE 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEE 120

Qy     121 TPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVD 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVD 180

Qy     181 LLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL 240

Qy     241 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTTILEAKQAKNPASPYNLAYKYN 300
```

```
           |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db     241 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKRAKNPASPYNLAYKYN 300

Qy     301 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350


RESULT   2
Q9UG89
ID   Q9UG89      PRELIMINARY;      PRT;   330 AA.
AC   Q9UG89;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 37.0 KDA PROTEIN (FRAGMENT).
GN   DKFZP564O0582.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Poustka A., Klein M., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL049929; CAB43210.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   330 AA;  37004 MW;  58098DCDF5BBCFE2 CRC64;

  Query Match            94.6%;  Score 1667;  DB 4;  Length 330;
  Best Local Similarity  99.7%;  Pred. No. 5.3e-119;
  Matches  329;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 ILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVMV 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVMVMV 60

Qy      81 KGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKA 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKA 120

Qy     141 NSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRH 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 NSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRH 180

Qy     201 KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAV 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 KHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAV 240

Qy     261 VELVTVKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVFNMVLWIMIALALA 320
           ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db     241 VELVTVKSFDTSLIRKTRTILEAKRAKNPASPYNLAYKYNFEYSVVFNMVLWIMIALALA 300

Qy     321 VIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
           ||||||||||||||||||||||||||||||
Db     301 VIITSYNIWNMDPGYDSIIYRMTNQKIRMD 330


RESULT   3
Q9VHG4
ID   Q9VHG4      PRELIMINARY;      PRT;   320 AA.
AC   Q9VHG4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG8444 PROTEIN.
GN   CG8444.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003682; AAF54350.1; -.
DR   FlyBase; FBgn0037671; CG8444.
SQ   SEQUENCE   320 AA;  34421 MW;  401BCEAAA4FBA7E5 CRC64;

  Query Match            18.5%;  Score 327;  DB 5;  Length 320;
  Best Local Similarity  26.3%;  Pred. No. 4.9e-17;
  Matches   94;  Conservative   74;  Mismatches  142;  Indels   48;  Gaps   11;

Qy       1 MAVFVVLLALVAGV-LGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLS 59
           : |||:    :| :   ||::|  | :: |: ||  :    : ||   |||  :|  | :
Db       2 LRVFVIFSLFIAAINASGEFTVLNRPKAISFK-GNDALESHYVGDVLYASMGNAVSGDTN 60

Qy      60 WPGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSE 119
           | || : : |:  :  ::| |:|:  :     |:| :| |  :      |:  |::
Db      61 WNGLTINDPFNLAKGVILVHVQGIGHVT-TAGNVKTYELTGS---GTDASLNAL------ 110

Qy     120 ETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEV 179
                 :| : | |  :        ||     :: ::  | :    :: |    |: :
Db     111 ----AAELEAANEPVCDIN-----FEQFDDGVQAWKS-CFGDFEAPAAKPTKHLNPS--- 157

Qy     180 DLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKI 239
             |  :: | | ::  : |   |||:   |     :     :| : | :|| |    :|:|:
Db     158 --LHTADKQFLQEVGFINSAADHLAEMAKPSNVLMLRVSVDGVAKAHGEKSVAVEEANKL 215

Qy     240 LVDALQKFADDMYSLYGGNAVVELVTVKSFDTSL-IRKTRTILEAKQAK------NPASP 292
           |  |: :                       : || |: | :: |   : | :||     :  :|
Db     216 LSAAISRLL--------------AASQKSSDSVLFVQTTEKDVAASRAKRDTIAASTTNP 261
```

```
Qy     293 YNLAYKYNFEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD 350
           ||||  |   :| |:||::||  |:    |:::    | |   |||| ||||||||: :|: |
Db     262 YNLAVYYGSDYPVIFNIILWFMVVFGLSLLAICYAIAAMDPGRDSIIYRMTSTRIKKD 319
```

```
RESULT   4
Q9XTY9
ID    Q9XTY9      PRELIMINARY;      PRT;   324 AA.
AC    Q9XTY9;
DT    01-NOV-1999 (TrEMBLrel. 12, Created)
DT    01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT    01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE    R03E1.2 PROTEIN.
GN    R03E1.2.
OS    Caenorhabditis elegans.
OC    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC    Rhabditidae; Peloderinae; Caenorhabditis.
OX    NCBI_TaxID=6239;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Mcmurray A.;
RL    Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=94150718; PubMed=7906398;
RA    Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA    Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA    Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA    Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA    Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA    Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA    Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA    Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA    Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA    Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT    "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT    elegans.";
RL    Nature 368:32-38(1994).
DR    EMBL; Z92837; CAB07401.1; -.
SQ    SEQUENCE   324 AA;  34217 MW;  D0B3D186CF861C6E CRC64;

  Query Match             10.9%;  Score 192.5;  DB 5;  Length 324;
  Best Local Similarity   22.3%;  Pred. No. 8.6e-07;
  Matches   86;  Conservative   64;  Mismatches  134;  Indels  101;  Gaps   13;

Qy       4 FVVLL-ALVAGVLGNEFSILKSPGSVVERNGNWPIP-------GERIPDVAALSM-GFSV 54
           ||:|: ||||     :    |:  :| |:  | :  :  :           | |   ::| : ||||
Db       3 FVLLVSALVAACAASSLEIIAAPASLKFPDADATLKTTLISTLNENILGLSARQVDGFSV 62

Qy      55 KEDLSWPGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIH 114
            |            :| ||||  :| | | : | |  |     |    :        |     |
Db      63 SVD----------MFSRPRALAVVTVIGADNLNL-GGKTYSTKSDGVENVGFDQDMALIF 111

Qy     115 SLFSEETPV--------VLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLS 166
                |   |         :| ||  :| |                 | :|  ::  |: |
Db     112 GADRENVQVSAGGITGSMLALAAQQEAV-----------DATV-IKTKRSTL-------- 151

Qy     167 SLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRY 226
                                :  ||:  :: ::: :  :    |::  |:: :  :  ||   |
Db     152 -----------------VKELEAVYQLAAAI-KSAGVQMDNNADVFRVSITGLVGI---- 189

Qy     227 GEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELV--------------------T 265
             :| |   :|   :   |::    :  :  ||| ||:||:
Db     190 -TESAQKEEAVADVKAAIEALTASINNAYGGQAVIELLAFEAESGAAEKSTESREIPNHN 248

Qy     266 VKSFDTSLIRKTRTILEAKQAKNPASPYNLAYKYNFEYSVVFNMVLWIMIALALAVIITS 325
           ::  ||      |:     |     |   | |             :|  :| :|  ::: |  :|:|
Db     249 IQKRDTKSKTPTQLARENLQVTVPVSS---------DYPAIFAIFLGLVVILVVALIYIV 299

Qy     326 YNIWNMDPGYDSIIYRMTNQKIRMD 350
             : ::||   ||||||||   ::: |
Db     300 VGMASIDPEKDSIIYRMTTTRMKKD 324
```

```
RESULT   5
Q9ZSJ7
ID    Q9ZSJ7      PRELIMINARY;      PRT;   364 AA.
AC    Q9ZSJ7;
DT    01-MAY-1999 (TrEMBLrel. 10, Created)
DT    01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT    01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE    PUTATIVE TYPE 1 MEMBRANE PROTEIN (GENOMIC DNA, CHROMOSOME 3, P1
DE    CLONE:MUJ8).
GN    PMP.
OS    Arabidopsis thaliana (Mouse-ear cress).
OC    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC    Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC    Brassicales; Brassicaceae; Arabidopsis.
OX    NCBI_TaxID=3702;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Goo J.H., Park A.R., Park W.J., Park O.K.;
RT    "Selection of Arabidopsis genes encoding secreted and membrane
RT    proteins.";
RL    Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=COLUMBIA;
RA    Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL    Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN    [3]
RP    SEQUENCE FROM N.A.
RC    STRAIN=COLUMBIA;
RX    MEDLINE=20277480; PubMed=10819329;
RA    Nakamura Y.;
RT    "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT    features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT    clones.";
RL    DNA Res. 7:131-135(2000).
DR    EMBL; AF104329; AAD11797.1; -.
DR    EMBL; AB028621; BAB01357.1; -.
SQ    SEQUENCE   364 AA;  39133 MW;  9A5CBC8CFD73D3A9 CRC64;

  Query Match              6.9%;  Score 121.5;  DB 10;  Length 364;
  Best Local Similarity   22.0%;  Pred. No. 0.26;
  Matches   86;  Conservative   72;  Mismatches  154;  Indels   79;  Gaps   18;

Qy       3 VFVVLLALVAGVLGNEFSILKSPGSVVF----------RNGNWPIPGERIPDVAALS-- 49
           |||| | |        |          :: ||| |              |: :  :|   :   ||:|
Db       6 VFVVALLLTLNYRG------EASGSVFFIDGSNNQYLRPRSSSEALPMSPVEISAAVSAL 59

Qy      50 MGFSVKEDLSWPG------LAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVP 103
           :||:    |:  |        :    | | ||||  :: :  | : : |         |: | ||:
Db      60 LGFAPSATLTADGSSKLNKILKPNPFERPRAAFVLEIAGADDMLLETSPSHSF-LGNAIR 118

Qy     104 FSLDSVANSIHSLFSEETPVVLQLAPSEERVYMVGKANSVFEDLSVTLRQLRNRL----- 158
           |: |         :  :|   :|  :|  |  |  : :|   |   :         : |
Db     119 SSIKS------DSYKADT----ELPDNEVVVVSVNEPSSDVTDKDI--NDFASWLGGSYV 166

Qy     159 --FQENSVLSSLPLNSLSRNNEVDL-------LFLSELQVLHDISSLLSRHKHLAK--DH 207
              : :| | |:||    | | :|          |: | :   :|     :| :   |:   |
Db     167 AGAEPSSGLLSIPLAG-GANVEFNLEKEAERKFALNLLGLYQNIRQAVSVYDDLSHGIDR 225

Qy     208 SPDLYSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVK 267
            : :|        |:| : : ||:      :     :|:  | |   : : : : |      ::| |
Db     226 TAELTVGRFGGIDALAQEYGQG--MAKQGMDVLLSTLSKLFNLLETSHKG----QIVGVI 279

Qy     268 SFDTSLIRKTRTILEAKQAKNPA----------SPYNLAYKYNFEYSVVFNMVLWIM-IA 316
              |   : :::  :|       :::  |         :|      |   :|   : |:   |
Db     280 VLDERVNQESENLLNFGSSRSSARSMVEVEGIPSAAIIA-----EVILVRLTLAWLTGII 334

Qy     317 LALAVIITSYNIWNMDPGYDSIIYRMTNQKI 347
            | :| |:   |  : ||     |:::|   :| |:
Db     335 LLIATILGVYFLMNMPLTKDTLLY--SNVKL 363
```

```
RESULT   6
O74359
ID   O74359         PRELIMINARY;    PRT;   663 AA.
AC   O74359;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 75.5 KDA PROTEIN.
GN   SPBC28E12.02 OR SPAC9B6.13.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031517; CAA20649.1; -.
DR   EMBL; AL049769; CAB42374.1; -.
DR   InterPro; IPR000566; -.
DR   InterPro; IPR000958; -.
DR   Pfam; PF00013; KH-domain; 1.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR   SMART; SM00322; KH; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   663 AA;  75492 MW;  6898B18AC142A8EA CRC64;

  Query Match             6.5%;  Score 114;  DB 3;  Length 663;
  Best Local Similarity   21.5%;  Pred. No. 2.5;
  Matches   62;  Conservative   43;  Mismatches   92;  Indels   92;  Gaps   13;

Qy       13 GVLGNEF-SILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHR 71
            ||  :|  ||: | ||:| ||       ||:         | ||  : || |
Db      414 GVSSSELSSIVSSTGSIVETNG----IGEK--------MSFSPLKKLSIP---------- 451

Qy       72 P-----RATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIHSLFSEETPVVLQ 126
            |      : ::| | || | |    :  :  || || ::|  :   :    :  :|
Db      452 PTEFVAQIAIICMASGVEML-LKTNGIEYFGQENTVPIAMDKASKIFYKFGQSQWHQILL 510

Qy      127 LAPSEERVYMVGKANSVFEDLSVTLR-QLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLS 185
             ||:::: :: || |   :  :   | |:|                      |:||
Db      511 EAPTKDQDFISGKKNGKLDKVKQQCRFNLKNG---------------------DILFCP 548

Qy      186 ELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGL-------------------DEIGKR 225
             :  :  :                 |:|| || :                  :|| |:
Db      549 QSTSIFTV---------------DIYSDELERVIKGMNTMLLEFPAEMHFYVPEEIHKK 592

Qy      226 -YGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSL 273
             |  || :  :| | :: :|:  :  || | ::   |   | :|
Db      593 LIGFRGEQIQRVTK-LYNSYIEFSTTPFDCYGHNVLIR--TPSKFSENL 638

RESULT   7
Q22736
ID   Q22736        PRELIMINARY;    PRT;   408 AA.
AC   Q22736;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T24D5.2 PROTEIN.
GN   T24D5.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z68012; CAA92020.1; -.
SQ   SEQUENCE   408 AA;  46861 MW;  B0B171CA84197EB9 CRC64;

  Query Match             6.2%;  Score 109.5;  DB 5;  Length 408;
  Best Local Similarity   20.8%;  Pred. No. 2.6;
  Matches   43;  Conservative   47;  Mismatches   84;  Indels   33;  Gaps    7;

Qy       81 KGVNK-LALPPGSVIS------YPLENAVPFSLDSVANSIHSL-FSEETPVVLQLAPSEE 132
             | :|: |  | | |::      : |:|    :  ::    : : |||    :|  :  :
Db      103 KAMNQFLKGPDGEVLAKYFQFCHELKNTTYKNKYTIRLDAYKMTFSERDDSILIRSDGDS 162

Qy      133 RVYMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHD 192
             |:|    |::|  : :: : :||| |   |   | || : : |:          : |: :
Db      163 TVHMGISNNTLFRNRNMAIEELRNMLSQNIRVYKSLFVTNRSQQ-------MGEMDLSKC 215

Qy      193 ISSLLSRHKHLAKDHSPDLYSLELAGLDEIGK-------RYGE--DSEQFRDASKILVDA 243
            :|:: : :  :   |  |::  :          : |          ||    : || |:|: : :|
Db      216 VSTITAENLTVVMDSFPNIIEIVKLHRPMLNKFTLMCSGEYGNIMELEQVRNAANLKLDG 275

Qy      244 LQKFADDMYSLYGGNAVVELVTVKSFD 270
             | |  ||          :  :|  :|:
Db      276 LSKITDDQ---------LNSITATNFE 293

RESULT   8
Q9N598
ID   Q9N598      PRELIMINARY;      PRT;   251 AA.
AC   Q9N598;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Y110A7A.11 PROTEIN.
GN   Y110A7A.11.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Ozersky P., Wohldmann P., Carter T.;
RT   "The sequence of C. elegans cosmid Y110A7A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
```

```
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006708; AAF60428.1; -.
SQ   SEQUENCE   251 AA;  28808 MW;  3E19F7FDD9B3FE40 CRC64;

  Query Match             6.1%;  Score 107.5;  DB 5;  Length 251;
  Best Local Similarity   22.1%;  Pred. No. 1.7;
  Matches   66;  Conservative   48;  Mismatches   86;  Indels   99;  Gaps   15;

Qy      55 KEDLSWPGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVANSIH 114
              |||:|               | |    | :||:         | |:|    |    | |
Db      23 KEDMS-------------ANVSAATKILNKMI--------YTLQN------DETTNDIL 54

Qy     115 SLFSEETPVVLQL---------APSEERVYMVGKANSVFEDLSVTLRQLRNRLFQEN--- 162
               : :|   ||||           : :|||: :: |   || |: ||:   :|     |:
Db      55 HEYRQE---VLQLKLLAEAESKSSAEERLKVIEKIPRVFPDVQVTVADSKNDNEAESYDF 111

Qy     163 SVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEI 222
                |:  | :  |:      :: |:|:       ||| :|| |:| |
Db     112 EKESAAGLRATQRS-----IYRSDLR---KQLLSSNKHRAQDTS------------- 147

Qy     223 GKRYGEDSEQFRD--ASKILVDALQKFADDMYSLY--GGNAVVELVTVKSFDTSLIRKTR 278
                   || |   ::     : | ::|      |       ::     |  : |       ||      |
Db     148 -----EDQEFMKNELVEEELANSLATMARSFKTMMSAAGDVIKE-------DTE-----R 190

Qy     279 TILEAKQAKNPASPYNL--------AYKYNFEYSVVFNMVLWIMIALALAVIITSYNIW 329
                || ||:  :  :   :               |||  ::     | ::| ::| ::    ::    |:
Db     191 AILMAKEVDDNKTALGIQSERVERHAYKCGYD---CFKVMLIVLIFMSFVSMVLMMKIF 246

RESULT   9
Q9FH47
ID   Q9FH47      PRELIMINARY;      PRT;  1112 AA.
AC   Q9FH47;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   GB|AAD55616.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
DR   EMBL; AB022212; BAB08882.1; -.
SQ   SEQUENCE   1112 AA;  127094 MW;  7C11A579BB26DE85 CRC64;

  Query Match             6.1%;  Score 107;  DB 10;  Length 1112;
  Best Local Similarity   21.5%;  Pred. No. 19;
  Matches   74;  Conservative   62;  Mismatches  114;  Indels   94;  Gaps   15;

Qy      75 TVMVMVKGVNKL---ALPPGSVISYPLENAVPF-------SLDSVANSIHSLFSEETPVV 124
             | : |: |  :   || | ::   ||: : |         | | |  :| | |  |  |
Db     592 TTKLFVRAVRDVFEGALHP-TLDFLDLEDCLNFIREHKSLSDDKVLQAIDLLKSVVTQKV 650

Qy     125 LQLAPSEERVYMVGKAN----------SVFEDLSVTLRQLRNRLFQENSVLSSLPLNSLS 174
             | :   : :: ::  :           | |:: :  |: |:  |   |   : ::   : |
Db     651 LLM---DTKILLIDNSRISLLNNLTRLSAFDNRTYILQLLKPFLLNE---IVNMESKAKS 704

Qy     175 RNNEVDLLFLSELQVLH------------------------DISSLLSRH-KHLAKDH 207
               | |||   | : |                        |||||  :|  :
Db     705 DAAEADLLNELEKEKLQPKEKPQSKEKPQSKKRRDRTKKKPSTSISSLLDKTVEHKPEST 764

Qy     208 SPDLYSLE---------LAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGN 258
              || |  ::|              ||     ::   |:|||        | | |:  |  :  :  |   |
Db     765 SPSLRTVEEDSMEPEDALASETDMHSRHGEDP---------LSDHLESAAGEAAARY--N 813

Qy     259 AVVE-----LVTVKSFDTSLIRKTRTILEAKQAKNPASPYNL-------AYKYNFEYSVV 306
             : ::      |:::| |   |:: ::  | : :| : | :  ||          | :      || :
Db     814 SALDMTLKALLSIKIFKEDLMQNSQQIQDQRQEQVPTALQNLFTAVVSEAIQDEGVYSCL 873

Qy     307 FNMVL------WIMIALALAVIITSYNIWN--MDPGYDSIIYRM 342
              :  |             | :  |  |::     |:|      :|  :|:: |:
Db     874 LSDFLTSQEEFLSMSSDAAKVVVAVLNLWRCWKNPERESLVTRL 917

RESULT  10
Q93377
ID   Q93377      PRELIMINARY;      PRT;   458 AA.
AC   Q93377;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   C44H4.1 PROTEIN.
GN   C44H4.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smye R.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z79598; CAB01867.1; -.
DR   InterPro; IPR001611; -.
DR   InterPro; IPR003591; -.
DR   Pfam; PF00560; LRR; 10.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00369; LRR_TYP; 1.
SQ   SEQUENCE   458 AA;  51171 MW;  8A6810B119684152 CRC64;

  Query Match             6.0%;  Score 105;  DB 5;  Length 458;
  Best Local Similarity   22.5%;  Pred. No. 6.7;
  Matches   69;  Conservative   47;  Mismatches  116;  Indels   74;  Gaps   12;

Qy      18 EFSILKSPGSVVFRNGNWPIPGERIPDVAALSMGFSVKEDLSWPGLAVGNLFHRPRATVM 77
             :  ||  | :::               | : | | ||                    :|    : :
Db     149 QLKILSLPDNLI----------EYVQDNAFLS------------------YHSRDSLLK 179

Qy      78 VMVKGVNKLALPPGSVISYPLENAVPFSLD-SVANSIHSLFSEETPVVLQLAPSEERVYM 136
              : :    | |: |  :    |||    ||| :: : |  |    |   |:    |::
Db     180 LDLSANNLTAIHPTGLLG--LENLSQLSLDKNLLSEIPSQALENIPSLEDLSLGVNRIHT 237

Qy     137 VGKANSV----FEDLSVTLRQLR---NRLFQENSVLSSLPL--NSLSRNNEVDLLFLSEL 187
             : : ||:      : ||: : |:|    :  |  | :|| |  |  | |:  :     |  :  |
Db     238 ISR-NSLPLPNLKSLSLEVNQIRLIPSDSFSETPLLSYLYLGNNLLTSIDASMFLHIGGL 296

Qy     188 QVL-----HDISSLLSRHK-----------HLAKDHSPDLYSLELAGLDEIGKRYGEDSE 231
             :||       ||:|: :   |               ||   |:|  |   |||
```

```
Db     297 KVLSMSNNKDITSIQANGKLSVQQFKLCICILAFQHAPSLIRLELFDC------------ 344

Qy     232 QFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDTSLIRKTRTILEAKQAKNPAS 291
                           |:|    :|||        |    | : ::  |   |   | :      | |:      :   :
Db     345 ---SISRIEPKSLQKVQHIQVILLSRNQITQINAVDDFAFSQL-PMLTSLDLSSNRLESL 400

Qy     292 PYNLAY 297
           | |: |
Db     401 PSNVIY 406
```

```
RESULT 11
025885
ID   025885      PRELIMINARY;      PRT;   412 AA.
AC   025885;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 47.1 KDA PROTEIN.
GN   HP1327.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000634; AAD08377.1; -.
DR   TIGR; HP1327; -.
DR   InterPro; IPR003423; -.
DR   Pfam; PF02321; OEP; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   412 AA;  47115 MW;  1724D84A9BFA15DD CRC64;

  Query Match             5.9%;  Score 104;  DB 2;  Length 412;
  Best Local Similarity   22.0%;  Pred. No. 6.8;
  Matches   70;  Conservative   55;  Mismatches  101;  Indels   92;  Gaps   15;

Qy      48 LSMGFSVKEDLSWPGLAVGNLFH-RPRATVMVMVKGVNKLALPPGSVISYPLEN------ 100
           :|:| | | ||:    |     :      :  :: : |    :||:    |::    :||
Db      97 MSLGLSQKVDLNGKKLTQSQMIDLEKQKKILELKKTKQQLAI---SLMINGIENYKNQQE 153

Qy     101 --AVPFSLDSVANSIHSLFSEETPVVLQLAPSE-ERVYMVGKANSVFEDLSVTLRQLRNR 157
             : :: :: |:::         :| :: :|  |   :   :  | |:: | || :    :
Db     154 IELLKTAIKNLENTLYQANHSSSPNLIAIAKLEILKSQLEIKKNNLEEALSASHYSMGEL 213

Qy     158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELA 217
            |:|| :||   | |      | |              | ||:||:              | : :|
Db     214 AFKENELLSIAPKN-FEFNRE---------QELHNISA-------------TNYDIAIA 249

Qy     218 GLDEIGKRYGEDSEQFRDASK--ILVD----------ALQKFADDMYS--------LYGG 257
            |||        | |::    :|    | |              : | :   ||:|        :||
Db     250 RLDE------EKSQKDITLAKKSFLEDVNVTGVYYFRSKQYYNYDMFSIALSIPLPIYGK 303

Qy     258 NAVVELVTVKSFDT-------------------SLIRKTRTI---LEA-----KQAKNPA 290
            |   :||  |   ::                             |::|    |:    ||:    |  : |
Db     304 QA--KLVEQKKKESLVFKSEVENTKNKTHHLALKLLKKLETLQKNLESINKIIKQNEKIA 361

Qy     291 SPYNLAYKYNFEYSVVFN 308
                            |  |   |  | :|:   :|
Db     362 QIYALDLKSNGDYNAYYN 379
```

```
RESULT 12
068717
ID   068717      PRELIMINARY;      PRT;   732 AA.
AC   068717;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PROTEIN KINASE A HOMOLOG.
GN   YPKA.
OS   Yersinia pestis.
OG   Plasmid pCD1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM;
RA   Hu P., Elliott J., McCready P., Skowronski E., Garnes J.,
RA   Kobayashi A., Carrano A.V., Brubaker R., Garcia E.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5;
RX   MEDLINE=98427122; PubMed=9746557;
RA   Perry R.D., Straley S.C., Fetherston J.D., Rose D.J., Gregor J.,
RA   Blattner F.R.;
RT   "DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of
RT   Yersinia pestis KIM5.";
RL   Infect. Immun. 66:4611-4623(1998).
CC   -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AF053946; AAC62602.1; -.
DR   EMBL; AF074612; AAC69765.1; -.
DR   InterPro; IPR000719; -.
DR   InterPro; IPR002290; -.
DR   InterPro; IPR003547; -.
DR   PRINTS; PR01373; YERSSTKINASE.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   SMART; SM00220; S_TKc; 1.
KW   ATP-binding; Kinase; Plasmid; Serine/threonine-protein kinase;
KW   Transferase.
SQ   SEQUENCE   732 AA;  81732 MW;  1ECB0099EE2D3430 CRC64;

  Query Match             5.9%;  Score 103.5;  DB 2;  Length 732;
  Best Local Similarity   21.2%;  Pred. No. 18;
  Matches   73;  Conservative   51;  Mismatches  103;  Indels  117;  Gaps   17;

Qy      12 AGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSM---------GFSVKEDLSWPG 62
           |||: |:       ||:|||   :    ||  | |  | :            ||:  |    |
Db     264 AGVVHNDI----KPGNVVFDRAS----GE--PVVIDLGLHSRSGEQPKGFT--ESFKAPE 311

Qy      63 LAVGNLFHRPRATVMVMV------------------KGVNKLALPPGSVISYPLENAVP 103
           | ||||                                :|:  :   |  |:    ||  |
Db     312 LGVGNLGASEKSDVFLVVSTLLHCIEGFEKNPEIKPNQGLRFITSEPAHVMD---ENGYP 368

Qy     104 FSLDSVANSIHSLFSEETPVVLQLA------PSEERVYMVGKANSVFEDLSVTLRQLRNR 157
             :|  : : ::        :| ::          :| |::           :: |: :   : |::
Db     369 IHRPGIA-GVETAYTRFITDILGVSADSRPDSNEARLHEFLSDGTIDEESAKQI--LKDT 425

Qy     158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL------ 211
            |  | | ||:             |:  ::   : |  ::| ||   |     |       |:
Db     426 LTGEMSPLST-----------DVRRITP-KKLRELSDLLRTHLSSAATKQLDMGGVLSD 472

Qy     212 YSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDT 271
               |  ||:  :  |  | :|                                 :|||:
Db     473 LDTMLVALDKAEREGGVDKDQ-------------------------------LKSFN- 498

Qy     272 SLIRKTRTILE---------AKQAKNPASPYNLAYKYNFEYSVV 306
```

```
           ||| ||  ::|             | :    |||   :: ||  |:|
Db     499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539

RESULT  13
Q9RI12
ID   Q9RI12      PRELIMINARY;      PRT;   732 AA.
AC   Q9RI12;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE TARGETED EFFECTOR PROTEIN KINASE.
GN   YPCD1.72C.
OS   Yersinia pestis.
OG   Plasmid pCD1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 BIOVAR ORIENTALIS;
RA   Baker S.G., Mungall K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 BIOVAR ORIENTALIS;
RA   James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 BIOVAR ORIENTALIS;
RA   Karlyshev A.V., Wren B.W.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AL117189; CAB54949.1; -.
DR   InterPro; IPR000719; -.
DR   InterPro; IPR002290; -.
DR   InterPro; IPR003547; -.
DR   PRINTS; PR01373; YERSSTKINASE.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   SMART; SM00220; S_TKc; 1.
KW   ATP-binding; Kinase; Plasmid; Serine/threonine-protein kinase;
KW   Transferase.
SQ   SEQUENCE   732 AA;  81760 MW;  DEF803AAEE2F5BCD CRC64;

  Query Match             5.9%;  Score 103.5;  DB 2;  Length 732;
  Best Local Similarity   21.2%;  Pred. No. 18;
  Matches   73;  Conservative   51;  Mismatches  103;  Indels  117;  Gaps   17;

Qy      12 AGVLGNEFSILKSPGSVVFRNGNWPIPGERIPDVAALSM---------GFSVKEDLSWPG 62
           |||: |:        ||:|||    :     ||  | |  | :        ||:  |    |
Db     264 AGVVHNDI----KPGNVVFDRAS----GE--PVVIDLGLHSRSGEQPKGFT--ESFKAPE 311

Qy      63 LAVGNLFHRPRATVMVMV------------------KGVNKLALPPGSVISYPLENAVP 103
            | ||||     :: | ::|                   :|:  :    |  |:     ||  |
Db     312 LGVGNLGASEKSDVFLVVSTLLHCIEGFEKNPEIKPNQGLRFITSEPAHVMD---ENGYP 368

Qy     104 FSLDSVANSIHSLFSEETPVVLQLA----PSEERVYMVGKANSVFEDLSVTLRQLRNR 157
               :|  : : ::    :| ::       :| |::     :: |: :  :  |::
Db     369 IHRPGIA-GVETAYTRFITDILGVSADSRPDSNEARLHEFLSDGTIDEESAKQI--LKDT 425

Qy     158 LFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL------ 211
            |  | | ||:               |:   ::    : | ::|  ||   |     |      |:
Db     426 LTGEMSPLST-------------DVRRITP-KKLRELSDLLRTHLSSAATKQLDMGGVLSD 472

Qy     212 YSLELAGLDEIGKRYGEDSEQFRDASKILVDALQKFADDMYSLYGGNAVVELVTVKSFDT 271
                |  ||:  :  | | :|                                 :|||:
Db     473 LDTMLVALDKAEREGGVDKDQ-------------------------------LKSFN- 498

Qy     272 SLIRKTRTILE---------AKQAKNPASPYNLAYKYNFEYSVV 306
           ||| ||  ::|             | :    |||   :: ||  |:|
```

```
Db     499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539

RESULT  14
Q9SIX4
ID   Q9SIX4      PRELIMINARY;      PRT;   925 AA.
AC   Q9SIX4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE LRR RECEPTOR PROTEIN KINASE.
GN   AT2G16250.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL   Nature 402:761-768(1999).
DR   EMBL; AC007047; AAD22312.1; -.
DR   InterPro; IPR000719; -.
DR   InterPro; IPR001611; -.
DR   InterPro; IPR003592; -.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00560; LRR; 7.
DR   PRINTS; PR00019; LEURICHRPT.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   SMART; SM00370; LRR; 1.
KW   ATP-binding; Kinase; Receptor; Transferase.
SQ   SEQUENCE   925 AA;  101125 MW;  C363800D4ABC39D3 CRC64;

  Query Match             5.9%;  Score 103.5;  DB 10;  Length 925;
  Best Local Similarity   22.6%;  Pred. No. 26;
  Matches   86;  Conservative   43;  Mismatches  126;  Indels  125;  Gaps   16;

Qy       4 FVVLL--------ALVAGVLGNEFS----ILKSPGSVVFRNGNWPIPGERIPD------- 44
           ||:||          :| |    : ||     :|     |:  |  :|||  |:   |
Db      11 FVLLLLCLVLFFDCVVVGQTQSRFSEKLILLNLRSSLGLRGTDWPIKGDPCVDWRGIQCE 70

Qy      45 ---------------VAALSMGFSV------------KEDLSWPGLAVGNLFHRPRAT 75
                          :  |:  |||                   |: ||  :   |
Db      71 NGSIIGINISGFRRTRIGKLNPQFSVDPLRNLTRLSYFNASGLALPG-TIPEWF------ 123

Qy      76 VMVMVKGVNKLALPPGSVISYPLENAVPFSLDSVA---------NSIHSLFSEETPVVLQ 126
                 ||: |||    : |  :   |||:| ::             ||: ||       :|
Db     124 ------GVSLLALEVLDLSSCSVNGVVPFTLGNLTSLRTLNLSQNSLTSLVPSSLGQLLN 177

Qy     127 LAPSE-ERVYMVGKANSVFEDLS--VTLRQLRNRL-----------------FQENSVL 165
           |:  :  |     |     |  |  :||    ||                    |   ||
Db     178 LSQLDLSRNSFTGVLPQSFSSLKNLLTLDVSSNYLTGPIPPGLGALSKLIHLNFSSNSFS 237

Qy     166 SSLP--------------LNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPD 210
           | :|                 :||||  :   :|  ||:||::      :|||          |
Db     238 SPIPSELGDLVNLVDFDLSINSLSGSVPQELRKLSKLQLMAIGDNLLSGTLPV------D 291

Qy     211 LYSLELAGLDEIGKRYGEDSEQFRDAS------KILVDALQKFADDM-YSLYGGNAVVEL 263
           |:| | : |   :   |     |           :||  |   |    :  ||  | : : |:
Db     292 LFSAE-SQLQTLVLRENGFSGSLPDVCWSLPKLRILDIAKNNFTGLLPYSSYDSDQIAEM 350

Qy     264 VTVKSFD-----TSLIRKTR 278
            | : |          | ::|: |
```

```
Db    351 VDISSNTFYGELTPILRRFR 370


RESULT 15
Q9ZDN3
ID   Q9ZDN3       PRELIMINARY;      PRT;   470 AA.
AC   Q9ZDN3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GUANOSINE PENTAPHOSPHATE PHOSPHOHYDROLASE (GPPA).
GN   RP294.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
DR   EMBL; AJ235271; CAA14755.1; -.
SQ   SEQUENCE   470 AA;  53780 MW;  74F7E97599D02903 CRC64;


  Query Match             5.8%;  Score 103;  DB 2;  Length 470;
  Best Local Similarity   21.7%;  Pred. No. 10;
  Matches   65;  Conservative   56;  Mismatches  123;  Indels   56;  Gaps   12;

Qy     45 VAALSMGPSVKEDLSWPGLAVCNLFHRPRATVMVMVKGVNKL-ALPPGSVISYPLENAVP 103
          :: :|  | : ||    | : :::             | || :|| |: |
Db    117 ISGISDAFGIVADLGGGSLELAHIYN----------KKVGKLKSLPLGTKI--------- 157

Qy    104 FSLDSVANS-------IHSLFSEETPVVLQLAPSEERVYMVGKA-----NSVFEDLSVTL 151
             :|||       |  :  ||  |             :|::| |         | ::  |
Db    158 -----IANSNFSDVGLITKMLEEEPGVA-----HYPNLYLIGGALRLMSRIYMESINYPL 207

Qy    152 RQLRNRLFQENSVLSSLPLNSLSRNNEVDLLFLSELQVLHDISSLLSRHKHLAKDHSPDL 211
          : | |  |: | |   | |  ||: ::: | :  :  : ::   ::   | : |  ||:
Db    208 KNLHN--FEINRVEPELYLEKLSQIDKLKLSYYEQKAINYNAVLVI---KAMIKVFSPEK 262

Qy    212 YSLELAGLDEIGKRYGEDSEQFRDASKIL----VDALQKFADDMYSLYGGNAVVELVTVK 267
          :   || | | |:  ||  : :  | :   |  |  |  ::  :          :: : :
Db    263 IIISNYGLKE-GVRF--DSLPYHETEKDIIYERVKRLVNFDRNICKIEKYIEALQYLLIN 319

Qy    268 SFDTSLIRKTRTILEAKQAKN--PASPYNLAYKYNFEYSVVFNMVLWIMIALALAVIITS 325
          |  |:|| |: |:  ||     |  ::    : |:   :|: :|| |  |:
Db    320 SDATTLIIIELAIMLAQYNANIDKTLRANFVSEFILSSDIPFSHRQRLMLGIALTVTYTA 379


Search completed: July  3, 2001, 16:20:24
Job time: 156 sec
```